AN SVD APPROACH TO IDENTIFYING META-STABLE STATES OF MARKOV CHAINS*

DAVID FRITZSCHE†, VOLKER MEHRMANN‡, DANIEL B. SZYLD†, AND ELENA VIRNIK‡

Abstract. Being one of the key tools in conformation dynamics, the identification of meta-stable states of Markov chains has been subject to extensive research in recent years, especially when the Markov chains represent energy states of biomolecules. Some previous work on this topic involved the computation of the eigenvalue cluster close to one, as well as the corresponding eigenvectors and the stationary probability distribution of the associated stochastic matrix. Later, since the eigenvalue cluster algorithm turned out to be non-robust, an optimisation approach was developed. As a possible less costly alternative, we present an SVD approach to identifying meta-stable states of a stochastic matrix, where we only need the second largest singular vector. We outline some theoretical background and discuss the advantages of this strategy. Some simulated and real numerical examples illustrate the effectiveness of the proposed algorithm.

Key words. Markov chains; conformation dynamics; singular value decomposition

AMS subject classifications. 15A18, 15A51, 60J10, 60J20, 65F15

1. Introduction. The research for this paper has been motivated by the work on conformation dynamics, or more specifically, on the identification of meta-stable conformations of biomolecules done by Deuflhard, Schütte et al., see, e.g., [5], [6] and the references therein. This problem arises for instance in drug design, where it is important to study different conformations of the drug molecule in order to optimise its shape for best possible binding properties with respect to the target molecule [15]. Different conformations, also called aggregates or meta-stable states of a molecule are sets of states such that the transition within the set is very probable whereas the transition between these sets only rarely occurs.

The approach to identify meta-stable conformations of biomolecules presented in [5] involves the computation of the eigenvalue cluster close to one, the so-called Perron cluster, as well as the corresponding eigenvectors. The number of eigenvalues in the cluster, then, represents the number of different meta-stable states. Since this algorithm turned out to be non-robust, an optimasation approach was developed in [6]. In both approaches the Markov chain is assumed to be reversible in order to exploit the fact that the transition matrix is then symmetric with respect to an inner product, which uses the stationary distribution of the Markov chain. The main drawbacks of these approaches are firstly, that the identification of the Perron cluster may be difficult or even impossible if the transition matrix of the Markov chain has no significant spectral gaps; and secondly, the calculation of the stationary distribution, although usually well conditioned [8], [19], may be costly and badly conditioned if the Perron cluster contains many eigenvalues very close to 1; see, e.g., [18].

In this paper, we present a different approach to identifying meta-stable states of a Markov chain: we find a permutation of a given stochastic transition matrix of

*this version dated 25th September 2006

†Temple University, Department of Mathematics, 1805 North Broad Street, Philadelphia, PA 19122-6094, USA, {fritzsche,szyld}@temple.edu

‡Technische Universität Berlin, Institut für Mathematik, Straße des 17. Juni 136, 10623 Berlin, Germany, {mehrmann,virnik}@math.tu-berlin.de

a Markov chain, such that the resulting matrix is block diagonally dominant. The diagonal blocks, then, represent the different meta-stable states. In our method we do not need to know the number of meta-stable states in advance but instead it is calculated in the process. Hence, the functionality of the algorithm does not depend on a significant gap in the spectrum of the transition matrix of the Markov chain. Furthermore, instead of calculating many eigenvectors or employing costly optimisation procedures, we calculate only two singular vectors that correspond to the two largest singular values. This allows us to use iterative procedures such as Lanczos or Arnoldi iteration for computations [3]. Since we are dealing with singular vectors instead of eigenvectors, we do not need the stationary distribution to symmetrise the problem, since singular vectors are orthogonal.

The basic idea of our algorithm is to calculate the singular vector that corresponds to the second largest singular value, sort its entries and apply the permutation thus obtained to the transition matrix. This idea is based on an observation due to I. Slapnicar [16]. Our strategy partly reflects well studied ideas from the computer science literature. In graph partitioning, the Fiedler vector, which is the eigenvector corresponding to the second smallest eigenvalue of a Laplacian matrix plays an important role, see, e.g., [9], [14] for the basics and [1], [17] for further reading. Ideas of using the singular value decomposition for graph clustering can be found, e.g., in [7], [20] or in the case of the seriation and the consecutive ones problem, e.g., in [2].

Our paper is organised as follows. In Section 2 we introduce the notation and some well-known definitions and theorems that we will use throughout the paper. In Section 3, we formulate some theoretical results for uncoupled Markov chains followed by Section 4, where we translate these results to the nearly uncoupled case. In Section 5, we describe our algorithm in detail. Finally, in Section 6, we present some constructed and some real numerical examples that illustrate the functionality of our method.

2. Preliminaries. We call a vector $v \in \mathbb{R}^n$ *positive* and we write $v > 0$ if all entries v_i are positive. A matrix $T \in \mathbb{R}^{n\times n}$, $T = [t_{ij}]_{i,j=1,\dots,n}$ is called *positive (non-negative)* and we write $T > 0$ ($T \geq 0$) if all entries t_{ij} are positive (non-negative). The matrix T is called *reducible* if there exists a permutation matrix $P \in \mathbb{R}^{n\times n}$, such that $PTP^T = \begin{bmatrix} T_{11} & 0 \\ T_{21} & T_{22} \end{bmatrix}$, where T_{11}, T_{22} are square. Otherwise it is called *irreducible*. We call the matrix T *(strictly) diagonally dominant* if $|t_{ii}| > \sum_{\substack{j=1 \\ j\neq i}}^{n} |t_{ij}|$ for all $i = 1, \dots, n$. We denote by $\mathbb{1}$ the vector of all ones $(1, \dots, 1)^T$.

A scalar $\lambda \in \mathbb{R}$ is called an *eigenvalue* of the matrix $T \in \mathbb{R}^{n\times n}$ if a vector $v \in \mathbb{R}^n$, $v \neq 0$ exists, such that $Tv = \lambda v$. The vector v is called a *(right) eigenvector* of T associated with λ. Accordingly, a vector $w \in \mathbb{R}^n$, $w \neq 0$ with $w^T T = \lambda w^T$ is called a *(left) eigenvector* of T. Let $T \in \mathbb{R}^{n\times n}$ have the eigenvalues λ_i, $i = 1, \dots, n$. We call $\rho(T) = \max_{1\leq i\leq n} |\lambda_i|$ the *spectral radius of* T.

A process is called *finite homogeneous Markov chain* if it has n states $s_1, \dots, s_n$ and the transition probability $P[s_i \rightsquigarrow s_j] =: t_{ij}$ is time-independent. The matrix $T = [t_{ij}]_{i,j=1,\dots,n}$ satisfies $t_{ij} \geq 0$ and $\sum_{j=1}^{n} t_{ij} = 1$ for $i, j = 1, \dots, n$, i.e., it is *(row) stochastic* and it is called the *transition matrix* of a Markov chain. We denote by $x^k = [x_i^k]$ the *probability distribution vector*, where x_i^k is the probability that the system is in state s_i after k steps. We have, $x_i^k \geq 0$ and $\sum_{i=1}^{n} x_i^k = 1$ for each k. A distribution vector x is said to be *stationary* if $x^T T = x^T$. A matrix A is called *block*

stochastic if A is block-diagonal, i.e.,

$$A = diag(A_1, \ldots, A_m), \tag{2.1}$$

and the matrices $A_i \in \mathbb{R}^{n_i \times n_i}$, $i = 1, \ldots, m$, are (row) stochastic matrices. For every block A_i, we define sets S_i of n_i indices corresponding to the bock A_i. We have $\bigcup_{i=1}^{n} S_i = \{1, \ldots, n\}$ and $S_i \cap S_j = \emptyset$ for $i \neq j$. We define by $A(S_i, S_j)$ the subblock of A that contains entries a_{ij}, where $i \in S_i, j \in S_j$.

The (adjacency) graph of a matrix $A = (a_{ij})_{i,j=1,\ldots,n}$ is defined by letting the vertices represent the unknowns. There is an edge from node v_i to node v_j whenever $a_{ij} \neq 0$. We call a graph and, hence, the corresponding matrix simply connected if for all $i, j \in \{1, \ldots, n\}$ there exists a path from node i to node j or from node j to node i.

The well-known Perron-Frobenius Theorem, see, e.g., [4, p. 27], guarantees the existence and uniqueness of a stationary distribution.

THEOREM 2.1 (Perron-Frobenius Theorem). *Let $T \geq 0$ be irreducible with spectral radius $\rho(T)$. Then $\rho(T)$ is a simple eigenvalue and T has a positive left and right eigenvector corresponding to $\rho(T)$. A positive eigenvector x of a non-negative matrix T corresponds to $\rho(T)$.*

In this paper we apply the singular value decomposition to identify metastable states of a Markov chain. The following well-known theorem, see, e.g., [11, p. 70], states the existence of a singular value decomposition.

THEOREM 2.2 (SVD). *Let $A \in \mathbb{R}^{n \times n}$. Then, there exist orthogonal matrices $U = [u_1, \ldots, u_n] \in \mathbb{R}^{n \times n}$ and $V = [v1, \ldots, v_n] \in \mathbb{R}^{n \times n}$ such that*

$$A = U \Sigma V^T, \tag{2.2}$$

where $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$ and $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0$. We call $\sigma_1, \ldots, \sigma_n$ *singular values*, $u_1, \ldots, u_n$ *left singular vectors* and $v_1, \ldots, v_n$ *right singular vectors* of A. Singular values with multiplicity one are called *simple*.

COROLLARY 2.3. *For any orthogonal matrices U, V, the matrices B and UBV have the same singular values.*

3. Uncoupled Markov chains and SVD. In this section we formulate the theoretical basis for the sign structure approach that we use in our algorithm. In Lemma 3.1 we show how permutations of a block-stochastic matrix affect the SVD. In Theorem 3.2, following the lines of [5, Lemma 2.5], we show an important sign structure property for singular vectors. Subsequently, we explain how we use this property in our approach and state the advantages of this strategy.

LEMMA 3.1. *Let A be a block-stochastic matrix of the form (2.1) and let $B = PAP^T$, where P is any permutation matrix. Let*

$$B = U \Sigma V^T$$

be a singular value decomposition of A as in (2.2). Then,

$$A = \tilde{U} \Sigma \tilde{V}^T, \tag{3.1}$$

where $\tilde{U} = P^T U$ and $\tilde{V}^T = V^T P$, is a singular value decomposition of A. The matrices A and B have the same singular values.

Proof. Since $PP^T = I$ for any permutation matrix P, the matrices $\tilde{U} = P^T U$ and $\tilde{V}^T = V^T P$ are orthogonal. ∎

In the following we present an approach to obtain a permutation matrix $\tilde{P}$ that yields $\tilde{A} = \tilde{P}^T B \tilde{P}$ where $\tilde{A}$ is block-diagonal and reveals the hidden block-structure of B. We will determine such a $\tilde{P}$ by means of the singular value decomposition.

THEOREM 3.2. *Let A be a block-stochastic matrix of the form (2.1) with m diagonal blocks denoted by $A_1, \ldots, A_m$. Let*

$$A = \tilde{U} \Sigma \tilde{V}^T$$

be a singular value decomposition of A as in (2.2) and let $\tilde{u}_1, \ldots, \tilde{u}_m$ be the m left singular vectors corresponding to the largest singular value of each of the blocks $A_1, \ldots, A_m$, respectively. Associate with every state s_i its sign structure

$$\operatorname{sign}(s_i) := \left[\operatorname{sgn}(\tilde{u}_1)_i, \ldots, \operatorname{sgn}(\tilde{u}_m)_i\right], \tag{3.2}$$

where

$$\begin{aligned} \operatorname{sgn} : \mathbb{R} &\longrightarrow \{-1, 0, 1\} \\ x &\longmapsto \begin{cases} 1, & x > 0 \\ 0, & x = 0 \\ -1 & x < 0 \end{cases} . \end{aligned}$$

Then,

i) states that belong to the same block of A exhibit the same sign structure, i.e., for any A_j and all $k, l \in S_j$, we have $\operatorname{sign}(s_k) = sign(s_l)$;

ii) states that belong to different blocks of A exhibit different sign structure, i.e., for any A_i, A_j with $i \neq j$ and all $k \in S_i, l \in S_j$ we have $\operatorname{sign}(s_k) \neq sign(s_l)$.

Proof. i) Left singular vectors of a matrix A can be obtained by finding the eigenvectors of AA^T, since from (2.2) we get $AA^T = \tilde{U} \Sigma^2 \tilde{U}^T$, see, e.g., [11]. Note that the singular values of A are the square roots of the eigenvalues of AA^T.

Since we have assumed A to have m blocks, the graph of each A_i is simply connected. Therefore, the matrix product $A_i A_i^T$ is irreducible and we have $A_i A_i^T \geq 0$. Hence, by the Perron-Frobenius Theorem 2.1 we have that $\rho(A_i A_i^T)$ is a simple eigenvalue and the corresponding right eigenvector $\hat{u}_i$ is strictly positive. Thus, the vector

$$\tilde{u}_i = [0, \ldots, 0, \hat{u}_i^T, 0, \ldots, 0]^T \tag{3.3}$$

is an eigenvector of AA^T corresponding to the largest eigenvalue of the block $A_i A_i^T$, i.e., it is a left singular vector corresponding to the largest singular value of the block A_i. This implies that states that belong to the same block exhibit the same sign structure.

ii) Since by part i) all states that belong to the same block have the same sign structure, without loss of generality, we can assume that every block consists of only one state. Then, since $\tilde{U} = [\tilde{u}_1, \ldots, \tilde{u}_m] \in \mathbb{R}^{m \times m}$ is orthogonal, the rows of $\tilde{U}$ are also orthogonal and, hence, no two vectors can have the same sign structure. ∎

Note, that the same results can be obtained for the right singular vectors by considering the matrix $A^T A$ instead of AA^T.

To illustrate the sign structure property established in Theorem 3.2 we consider the following example.

EXAMPLE 3.3. *Consider a block diagonal transition matrix of a Markov chain with three blocks of sizes 2, 3, 2. Then, the three largest singular vectors v_1, v_2, v_3 corresponding to each of the blocks are linear combinations of the vectors $\tilde{u}_i$ in (3.3). We have that the vectors $\tilde{u}_i$*

$$\begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \\ s_6 \\ s_7 \end{array} \begin{array}{c} \tilde{u}_1 \\ \left[\begin{array}{c} + \\ + \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{array}\right. \end{array} \begin{array}{c} \tilde{u}_2 \\ \left[\begin{array}{c} 0 \\ 0 \\ + \\ + \\ + \\ 0 \\ 0 \end{array}\right. \end{array} \begin{array}{c} \tilde{u}_3 \\ \left[\begin{array}{c} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ + \\ + \end{array}\right. \end{array}$$

are non-negative on the block they correspond to and zero elsewhere. A possible linear combination for the orthogonal vectors v_i could lead to the following sign structure.

$$\begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \\ s_6 \\ s_7 \end{array} \begin{array}{c} v_1 \\ \left[\begin{array}{c} + \\ + \\ + \\ + \\ + \\ + \\ + \end{array}\right. \end{array} \begin{array}{c} v_2 \\ \left[\begin{array}{c} + \\ + \\ - \\ - \\ - \\ - \\ - \end{array}\right. \end{array} \begin{array}{c} v_3 \\ \left[\begin{array}{c} - \\ - \\ - \\ - \\ - \\ + \\ + \end{array}\right. \end{array} .$$

Here, the states s_1, s_2 belong to the first block and have the sign structure $(+,+,-)$, the states s_3, s_4, s_5 belong to the second block and have the sign structure $(+,-,-)$ and the states s_6, s_7 belong to the third block and have the sign structure $(+,-,+)$.

The idea to sort the singular vector corresponding to the second largest singular value and to apply the resulting permutation to the matrix is due to an observation by Slapnicar [16]. This method always works for matrices with only two blocks, see Section 5 for an Example, and usually works for matrices with a few more blocks. For larger matrices having more blocks, however, this simple approach is not sufficient to reveal the block structure.

By using the sign structure properties established in Theorems 3.1 and 3.2 we modify this idea into a recursive bisectioning algorithm that is suitable for large matrices with any number of blocks. The main strategy is to identify two blocks in each step and apply the sorting procedure recursively to each of the blocks. The details of the algorithm are presented in Section 5.

The advantages of this approach as opposed to the eigenvalue approach presented in [5] are the following:

- instead of computing all eigenvectors corresponding to the eigenvalue 1, we only calculate two singular vectors;
- singular vectors are orthogonal, hence we do not need the stationary distribution for the transition matrix to be symmetric in an inner product;
- we do not need to know the number of blocks in advance. Instead, we only set a tolerance threshold for the size of the entries in the off-diagonal blocks. The number of identified blocks, then, reflects the given tolerance, see Section 4;

- the approach makes use of combinatorial aspects of the problem.
- it is less costly than an optimisation approach.

4. Nearly uncoupled Markov chains. In the previous section we have considered uncoupled Markov chains. In applications, due to perturbations, noise and actual weak couplings between aggregates, the Markov chains are nearly uncoupled. Such a matrix B, consisting of m nearly uncoupled blocks, can be transformed by a permutation matrix P to

$$PBP^T = A + E = \begin{bmatrix} A_{11} & E_{12} & \dots & E_{1m} \\ E_{21} & A_{22} & \dots & E_{2m} \\ \vdots & \vdots & & \vdots \\ E_{m1} & E_{m2} & \vdots & A_{mm} \end{bmatrix}, \tag{4.1}$$

where the elements of each E_{ij} are small. In this case, we are looking for some permutation matrix $\tilde{P}$, possibly different from the matrix P in 4.1, that permutes B into a block diagonally-dominant matrix of the form 4.1. In order to define diagonal dominance for blocks, we need to introduce a measure for the smallness of the off-diagonal blocks or, equivalently, a measure for the largeness of the diagonal blocks.

For this purpose, following [5, Definitions 2.3, 2.4], we first define a norm that is more general and then use special cases of it.

Let $S_k, S_l \subseteq \{1, \dots, n\}$ be sets of indices. In the following, we denote by $B_{kl} = B(S_k, S_l)$ the sub-block of B corresponding to the index sets S_k, S_l. For simplicity, for any k, we write B_k for the diagonal block B_{kk}.

DEFINITION 4.1 (Conditional transition probability). *Let $B = [b_{ij}] \in \mathbb{R}^{n\times n}$ be a stochastic matrix. Let $v = [v_1, \dots, v_n]^T$ be a positive vector with $\sum_{i=1}^n v_i = 1$. Let $S_k, S_l \subseteq \{1, \dots, n\}$ be sets of indices with $S_k \cap S_l = \emptyset$ and let $B_k = B(S_k, S_k), B_l = B(S_l, S_l)$ be the corresponding blocks. Then, the conditional transition probability from B_k to B_l is given by*

$$\omega_v(B_k, B_l) = \frac{\sum_{i\in S_k, j\in S_l} v_i b_{ij}}{\sum_{i\in S_k} v_i}.$$

DEFINITION 4.2 (v-Norm). *For any vector $v > 0$, we define the v-norm of a non-negative matrix (block) B_{kl} by*

$$\|B_{kl}\|_v := \omega_v(B_k, B_l).$$

DEFINITION 4.3 (Coupling matrix). *Let $S_1, \dots, S_m \subseteq \{1, \dots, n\}$ be sets of indices such that $\bigcup_{i=1}^n S_i = \{1, \dots, n\}$ and $S_i \cap S_j = \emptyset$, for all $i \neq j$. Let $B_k = B(S_k, S_k)$, $k = 1, \dots, m$, be the diagonal blocks of the corresponding block decomposition of B. The coupling matrix of the decomposition is given by the stochastic matrix W_v defined by*

$$(W_v)_{kl} = \omega_v(B_k, B_l),$$

for $k, l = 1, \dots, m$. In [5] and [6] the vector v is taken to be the stationary distribution of the Markov chain, i.e., $v = \pi$, where $\pi^T B = \pi^T$ and $\pi^T \mathbb{1} = 1$. Hence, the norm used in [5] and [6] is called the π-norm.

DEFINITION 4.4 (𝟙-norm). *We set* $v = [1, \ldots, 1]^T =: \mathbb{1}^T$ *and obtain for the norm of a matrix (block)* B_{kl}

$$\|B_{kl}\|_{\mathbb{1}} = \frac{1}{n_k} \sum_{i \in S_k, j \in S_l} b_{ij}. \tag{4.2}$$

We call this norm the $\mathbb{1}$*-norm.* Note that the $\mathbb{1}$-norm is simply the average row sum of a matrix (block).

We discuss the difference of the $\mathbb{1}$-norm to the norm used in [5] and [6] in Section 6. The advantage of our choice is clearly that we avoid calculating the stationary distribution of the Markov chain, which, although usually well conditioned [8], [19], may be costly and badly conditioned if the Perron cluster contains many eigenvalues very close to 1, see, e.g., [18]. We claim to obtain the same blocks with both norms only with different coupling matrices. The following lemma gives the equivalence factors of the two norms.

LEMMA 4.5. *Let* $B = [b_{ij}] \in \mathbb{R}^{n \times n}$ *be a stochastic matrix. Let* $S_k \subseteq \{1, \ldots, n\}$ *be a set of* n_k *indices and* $B_k = B(S_k, S_k)$ *the corresponding principal sub-block. Furthermore, let* $v = [v_1, \ldots, v_n]^T$ *be a positive vector and* $v_{\min}$, $v_{\max}$ *the minimum and maximum values of the entries in* $v(S_k)$*. Then, we have*

$$\|B_k\|_v \leq \frac{v_{\max}}{v_{\min}} \|B_k\|_{\mathbb{1}} \leq \frac{v_{\max}^2}{v_{\min}^2} \|B_k\|_v.$$

Proof. We have

$$\|B_k\|_v = \frac{\sum_{i,j \in S_k} v_i b_{ij}}{\sum_{i \in S_k} v_i}, \quad \|B_k\|_{\mathbb{1}} = \frac{1}{n_k} \sum_{i,j \in S_k} b_{ij}.$$

Since $v_{\min} \leq v_i \leq v_{\max}$ for all $i \in S_k$, we have that

$$\|B_k\|_v \leq \frac{\sum_{i,j \in S_k} v_{\max} b_{ij}}{\sum_{i \in S_k} v_{\min}} = \frac{v_{\max}}{v_{\min}} . \frac{1}{n_k} \sum_{i,j \in S_k} b_{ij} = \frac{v_{\max}}{v_{\min}} \|B_k\|_{\mathbb{1}}.$$

Similarly,

$$\|B_k\|_v \geq \frac{\sum_{i,j \in S_k} v_{\min} b_{ij}}{\sum_{i \in S_k} v_{\max}} = \frac{v_{\min}}{v_{\max}} \|B_k\|_{\mathbb{1}},$$

giving us the two constants of the equivalence of the norms. ∎

In the numerical examples in Section 6, we can see that for the diagonal blocks the π-norm is usually larger than the $\mathbb{1}$-norm. Yet, this is not always the case as the following example demonstrates.

EXAMPLE 4.6. *Consider the stochastic matrix*

$$B = \left[\begin{array}{cc|c} 0.1 & 0.9 & 0 \\ 0.1 & 0.8 & 0.1 \\ \hline 0.3 & 0.1 & 0.6 \end{array}\right].$$

The stationary distribution of B *is given by* $\pi = \begin{bmatrix} 0.1346 & 0.6923 & 0.1731 \end{bmatrix}$*. For the first two by two block* B_1 *we get* $\|B_1\|_\pi = 0.9163 < 0.95 = \|B_1\|_{\mathbb{1}}$.

We now use the $\mathbb{1}$-norm to introduce a measure for the largeness of diagonal blocks.

DEFINITION 4.7 (Nearly uncoupled). *We call two blocks nearly uncoupled if the $\mathbb{1}$-norm of each of the blocks is larger than a given threshold $thr = 1 - \delta$ for some small $\delta > 0$. We call a matrix $\tilde{B}$ nearly uncoupled if it consists of m diagonal blocks and the corresponding coupling matrix, see Definition 4.3, is diagonally dominant.*

Our algorithm determines the maximal number of blocks such that the coupling matrix is diagonally dominant.

In the previous section we have shown that singular vectors that correspond to the largest singular values of each of the blocks have a specific sign structure. States that belong to the same block exhibit the same sign structure and states that belong to different blocks exhibit different sign structures. Since our identification algorithm is based on this sign structure, we need to show that under certain conditions the assertions of Theorem 3.2 are still true under perturbations.

4.1. Perturbation theory. Consider the perturbed stochastic matrix $B = \hat{A} + \epsilon R$ for some $\epsilon > 0$. For sufficiently small real ϵ the matrix $T(\epsilon) = BB^T$ is a linear symmetric operator that can be written by [12, pp.63,120] as

$$T(\epsilon) = T + \epsilon T^{(1)} + \mathcal{O}(\epsilon^2), \tag{4.3}$$

where $T(0) = T = \hat{A}\hat{A}^T$ is the unperturbed operator and $T^{(1)} = \hat{A}R^T + R\hat{A}^T$ is a Lyapunov perturbation operator. The matrix-valued function $T(\epsilon)$ is for all real $\epsilon > 0$ a product of a stochastic matrix with its transpose, that is symmetric and non-negative. Note, that the perturbations here are also symmetric. According to [12, Section 6.2] for such a $T(\epsilon)$ there exists an orthonormal basis of eigenvectors $\varphi_k(\epsilon)$ that are analytic functions of ϵ. In particular, the eigenvectors $\varphi_k(\epsilon)$ depend smoothly on ϵ and admit a Taylor expansion

$$\varphi_k(\epsilon) = \varphi_k + \epsilon\varphi_k^{(1)} + \mathcal{O}(\epsilon^2), \tag{4.4}$$

where φ_k are the orthonormal eigenvectors of the unperturbed operator T, i.e., linear combinations of the vectors $\tilde{u}_i$ in (3.3). The following is a generalisation of [5, Theorem 3.1] and [6, Lemma 2.1] from eigenvectors to singular vectors.

THEOREM 4.8. *Let $T(\epsilon)$ as in (4.3) have the two largest eigenvalues $\lambda_1(\epsilon), \lambda_2(\epsilon)$. Let $T = T(0)$ as in (4.3) be the corresponding unperturbed operator with eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. Then the perturbed orthonormal eigenvectors $\varphi_1(\epsilon), \varphi_2(\epsilon)$ corresponding to the perturbed singular values $\lambda_1(\epsilon), \lambda_2(\epsilon)$ are of the form*

$$\varphi_k(\epsilon) = \sum_{j=1}^{2}(\alpha_{ij} + \epsilon\beta_{ij})\tilde{u}_j + \epsilon\sum_{j=3}^{n}\left\langle \varphi_j, \varphi_k^{(1)} \right\rangle \varphi_j + \mathcal{O}(\epsilon^2), \tag{4.5}$$

for $k = 1, 2$, where $\tilde{u}_j$ are the eigenvectors in (3.3) and α_{ij}, β_{ij} are suitable coefficients.

Proof. Let us consider the more general case of m eigenvectors. Let T have a permutation PTP^T that has m uncoupled irreducible blocks and let $\lambda_1, \ldots, \lambda_m$ be the largest eigenvalues corresponding to each of the blocks, i.e., the corresponding eigenvectors φ_k are linear combinations of the vectors in (3.3). For $k = 1, \ldots, m$ let P_k be the orthogonal projection onto the eigenspace of the eigenvalue λ_k. Then, by

[12, Sec. II.2.1], the perturbed projection $P_k(\epsilon)$ is analytic in ϵ and admits a Taylor expansion

$$P_k(\epsilon) = P_k + \epsilon P_k^{(1)} + \mathcal{O}(\epsilon^2),\ k = 1, \ldots, m,$$

By [12, Sec. II.2.1(2.14)], we obtain

$$P_k(\epsilon) = P_k + \epsilon \sum_{\substack{j \in \{1,\ldots,n\} \\ j \neq k}} \frac{1}{\lambda_k - \lambda_j}(P_k T^{(1)} P_j + P_j T^{(1)} P_k) + \mathcal{O}(\epsilon^2),\ k = 1, \ldots, m.$$

Consider now the eigenvalues $\lambda_1, \ldots, \lambda_m$, where $\lambda_k \neq \lambda_j$ for $k \neq j$, $j, k = 1, \ldots, m$. Let $P_{1,\ldots,m}$ be the orthogonal projection onto the eigenspace corresponding to $\lambda_1, \ldots, \lambda_m$. Then,

$$\begin{aligned} P_{1,\ldots,m}(\epsilon) &= \sum_{i=1}^{m} P_i(\epsilon) = \\ &= \sum_{i=1}^{m} P_i + \epsilon \sum_{i=1}^{m} \sum_{\substack{j \in \{1,\ldots,n\} \\ j \neq i}} \frac{1}{\lambda_i - \lambda_j}(P_i T^{(1)} P_j + P_j T^{(1)} P_i) + \mathcal{O}(\epsilon^2) = \\ &= P_{1,\ldots,m} + \epsilon \sum_{i=1}^{m} \sum_{j=m+1}^{n} \frac{1}{\lambda_i - \lambda_j}(P_i T^{(1)} P_j + P_j T^{(1)} P_i) + \mathcal{O}(\epsilon^2), \end{aligned} \tag{4.6}$$

since the terms for $j \leq m$ cancel out. For the corresponding eigenvectors $\varphi_1(\epsilon), \ldots, \varphi_m(\epsilon)$, we have that

$$\varphi_k(\epsilon) = P_{1,\ldots,m}(\epsilon)\varphi_k(\epsilon),\ k = 1, \ldots, m. \tag{4.7}$$

By plugging in (4.4) and (4.6) into the right hand side of (4.7), we obtain

$$\begin{aligned} \varphi_k(\epsilon) &= \varphi_k + \epsilon \varphi_k^{(1)} + \mathcal{O}(\epsilon^2) = \\ &= P_{1,\ldots,m}\varphi_k + \epsilon\Big(\sum_{j=m+1}^{n} \frac{1}{\lambda_k - \lambda_j} P_j T^{(1)} \varphi_k + P_{1,\ldots,m}\varphi_k^{(1)}\Big) + \mathcal{O}(\epsilon^2). \end{aligned}$$

Comparing the coefficients corresponding to ϵ, we get

$$(I - P_{1,\ldots,m})\varphi_k^{(1)} = \sum_{j=m+1}^{n} \frac{1}{\lambda_k - \lambda_j} P_j T^{(1)} \varphi_k.$$

Since $(I - P_{1,\ldots,m})$ is the orthogonal projection complementary to $P_{1,\ldots,m}$, which is the projection onto the eigenspace corresponding to the eigenvectors $\varphi_1 \ldots, \varphi_m$ we obtain

$$\varphi_k^{(1)} = \sum_{j=1}^{m} \tilde{\beta}_{ij}\varphi_k + \sum_{j=m+1}^{n} \frac{1}{\lambda_k - \lambda_j} P_j T^{(1)} \varphi_k. \tag{4.8}$$

with some coefficients $\tilde{\beta}_{ij} \in \mathbb{R}$. By inserting (4.8) into (4.4), we obtain

$$\begin{aligned} \varphi_k(\epsilon) &= \varphi_k + \epsilon \varphi_k^{(1)} + \mathcal{O}(\epsilon^2) = \\ &= \sum_{j=1}^{m} (\alpha_{ij} + \epsilon\beta_{ij})\tilde{u}_j + \epsilon \sum_{j=m+1}^{n} \frac{1}{\lambda_k - \lambda_j} P_j T^{(1)} \varphi_k + \mathcal{O}(\epsilon^2), \end{aligned} \tag{4.9}$$

with some coefficients $\alpha_{ij}, \beta_{ij} \in \mathbb{R}$ since the eigenvectors φ_k are linear combinations of the vectors $\tilde{u}_j$ in (3.3).

Following now the lines of the proof of [6, Lemma 2.1] we can rewrite the second summand in (4.9) as follows. First, for $k = 1, \ldots, m$, we expand the perturbed eigenvalues

$$\lambda_k(\epsilon) = \lambda_k + \epsilon\lambda_k^{(1)} + \mathcal{O}(\epsilon^2), \tag{4.10}$$

and rewrite the second summand as a projection in terms of the Euclidean scalar product $\langle , \rangle$

$$\sum_{j=m+1}^{n} \frac{1}{\lambda_k - \lambda_j} P_j T^{(1)} \varphi_k = \sum_{j=m+1}^{n} \frac{1}{\lambda_k - \lambda_j} \left\langle \varphi_j, T^{(1)}\varphi_k \right\rangle \varphi_j.$$

Now we need an expression for $T^{(1)}$. For $k = 1, \ldots, m$ we have

$$T(\epsilon)\varphi_k(\epsilon) = \lambda_k(\epsilon)\varphi_k(\epsilon).$$

We insert all expansions and obtain

$$(T + \epsilon T^{(1)} + \mathcal{O}(\epsilon^2))(\varphi_k + \epsilon\varphi_k^{(1)} + \mathcal{O}(\epsilon^2)) = (\lambda_k + \epsilon\lambda_k^{(1)} + \mathcal{O}(\epsilon^2))(\varphi_k + \epsilon\varphi_k^{(1)} + \mathcal{O}(\epsilon^2)).$$

The zero order coefficient comparison of the above equation yields

$$T\varphi_k = \lambda_k\varphi_k.$$

The first order coefficient comparison yields

$$T\varphi_k^{(1)} + T^{(1)}\varphi_k = \lambda_k\varphi_k^{(1)} + \lambda_k^{(1)}\varphi_k,$$

which transforms to

$$T^{(1)}\varphi_k = (\lambda_k I - T)\varphi_k^{(1)} + \lambda_k^{(1)}\varphi_k.$$

Finally, we can rewrite the scalar product expression.

$$\begin{aligned} \left\langle \varphi_j, T^{(1)}\varphi_k \right\rangle &= \left\langle \varphi_j, (\lambda_k I - T))\varphi_k^{(1)} + \lambda_k^{(1)}\varphi_k \right\rangle = \\ &= \left\langle \varphi_j, (\lambda_k I - T)\varphi_k^{(1)} \right\rangle + \underbrace{\lambda_k^{(1)} \langle \varphi_j, \varphi_k \rangle}_{=0}. \end{aligned}$$

The last term vanishes due to orthogonality of the unperturbed eigenvectors of a symmetric operator. For the first term, since T is symmetric, we obtain

$$\begin{aligned} \left\langle \varphi_j, (\lambda_k I - T)\varphi_k^{(1)} \right\rangle &= \left\langle (\lambda_k I - T)\varphi_j, \varphi_k^{(1)} \right\rangle = \\ &= (\lambda_k - \lambda_j) \left\langle \varphi_j, \varphi_k^{(1)} \right\rangle. \end{aligned}$$

Now, we can rewrite (4.9) as

$$\varphi_k(\epsilon) = \sum_{j=1}^{m} (\alpha_{ij} + \epsilon\beta_{ij})\tilde{u}_j + \epsilon \sum_{j=m+1}^{n} \left\langle \varphi_j, \varphi_k^{(1)} \right\rangle \varphi_j + \mathcal{O}(\epsilon^2),$$

which for $m = 2$ is the result in (4.5). ∎

Theorem 4.8 that the second singular vector has the following representation

$$\varphi_2(\epsilon) = (\alpha_{21} + \epsilon\beta_{21})\tilde{u}_1 + (\alpha_{22} + \epsilon\beta_{22})\tilde{u}_2 + \epsilon \sum_{j=m+1}^{n} \left\langle \varphi_j, \varphi_2^{(1)} \right\rangle \varphi_j + \mathcal{O}(\epsilon^2)$$

The first two terms do not spoil the sign structure as long as α_{ij}, β_{ij} have the same sign or in case of different sign ϵ is small enough such that $|\alpha_{i,j}| > \epsilon|\beta_{ij}|$ holds, respectively. The third term depends on the orthogonality of the first order perturbation of the second singular vector $\varphi_2^{(1)}$ with respect to the rest of the singular vectors. If it is close to orthogonal, this term will be close to zero.

5. The algorithm. In this section we propose an algorithm that finds a permutation of a stochastic matrix that permutes it into block diagonally dominant form 4.1 by recursively identifying diagonally dominant blocks. We first present an identification procedure in the case of two blocks. Then, we imbed this procedure into a recursive method that works for any number of blocks.

Consider the case of two nearly uncoupled blocks. Let $PBP^T = A + E$ be a matrix of the form (4.1) that consists of two blocks ($m = 2$), where P is not known.

Since B is not uncoupled, BB^T is irreducible and from the Perron-Frobenius Theorem 2.1 we know that B has a simple largest singular value σ_1 and a corresponding positive left singular vector u_1. Consider now the second largest singular value σ_2 and the corresponding left singular vector u_2.

Since singular vectors are orthogonal (or by the Perron-Frobenius theorem), u_2 must have a change in sign, i.e., there exist two indeces i and j such that $(u_2)_i(u_2)_j < 0$. From Theorem 3.2 we know that states with different sign must belong to different blocks. We sort the second singular vector, e.g., in increasing order and use this permutation to permute the matrix B. Then, we split the permuted matrix $\tilde{B} = \tilde{P}B\tilde{P}^T$ such that the first block $\tilde{B}_1$ is of size equal to the number of negative values in u_2 and the second block $\tilde{B}_2$ is of size equal to the number of positive values in u_2. By the proof to [10, Theorem 2.3], we know that in our case the second singular value σ_2 is smaller or equal to the largest singular values of $\tilde{B}_1$ and $\tilde{B}_2$, respectively. Hence, the results of Theorem 3.2 and Section 4 on sign structure apply to the vector u_2. We obtain a matrix $\tilde{B}$ that reveals the hidden block structure of B, i.e., $\tilde{B}$ has the same diagonal block structure as A up to a permutation of the blocks and the entries within a block.

The following example demonstrates what we described above.

EXAMPLE 5.1. *Consider the row stochastic matrix*

$$A = \begin{bmatrix} 0.2000 & 0.8000 & 0 & 0 & 0 \\ 0.4000 & 0.6000 & 0 & 0 & 0 \\ 0 & 0 & 0.3000 & 0.3000 & 0.4000 \\ 0 & 0 & 0.2000 & 0.2000 & 0.6000 \\ 0 & 0 & 0.1000 & 0.1000 & 0.8000 \end{bmatrix}.$$

The singular vectors of A are given by

$$U = \begin{bmatrix} 0 & -0.7555 & 0 & -0.6552 & 0 \\ 0 & -0.6552 & 0 & 0.7555 & 0 \\ -0.4590 & 0 & 0.7891 & 0 & 0.4082 \\ -0.5702 & 0 & 0.0908 & 0 & -0.8165 \\ -0.6813 & 0 & -0.6076 & 0 & 0.4082 \end{bmatrix}.$$

We can see that the singular vectors are only non-zero on the one block they correspond to.

Now consider the perturbed matrix $A + E$, where

$$E = \begin{bmatrix} -0.0345 & -0.1128 & 0.0887 & 0.0393 & 0.0192 \\ -0.0592 & -0.1041 & 0.0667 & 0.0421 & 0.0544 \\ 0.0428 & 0.0015 & -0.0473 & 0.0154 & -0.0125 \\ 0.0252 & 0.0636 & -0.0212 & -0.0112 & -0.0564 \\ 0.0720 & 0.0799 & 0.0252 & -0.0656 & -0.1116 \end{bmatrix}.$$

such that the matrix

$$\tilde{A} = A + E = \left[\begin{array}{cc|ccc} 0.1655 & 0.6872 & 0.0887 & 0.0393 & 0.0192 \\ 0.3408 & 0.4959 & 0.0667 & 0.0421 & 0.0544 \\ \hline 0.0428 & 0.0015 & 0.2527 & 0.3154 & 0.3875 \\ 0.0252 & 0.0636 & 0.1788 & 0.1888 & 0.5436 \\ 0.0720 & 0.0799 & 0.1252 & 0.0344 & 0.6884 \end{array}\right],$$

is row stochastic again. Here, the error matrix E has been obtained as follows. We choose a random matrix R of same size as E, where the entries that correspond to the diagonal blocks are uniformly distributed in $(-1, 1)$ and the entries that correspond to the off-diagonal blocks are uniformly distributed in $(0, 1)$. Then, we compute the matrix $A + \epsilon R$, where $\epsilon := 10^{-2}$, which is not stochastic any longer. Now, we scale each row by dividing each of its elements by the row sum and obtain the stochastic matrix $\tilde{A}$. We have $E = \tilde{A} - A$ and E has row sums zero.

Furthermore, consider a permuted matrix $B = P\tilde{A}P^T$, where P is a random symmetric permutation matrix. We obtain

$$B = \begin{bmatrix} 0.1655 & 0.0393 & 0.0887 & 0.0192 & 0.6872 \\ 0.0252 & 0.1888 & 0.1788 & 0.5436 & 0.0636 \\ 0.0428 & 0.3154 & 0.2527 & 0.3875 & 0.0015 \\ 0.0720 & 0.0344 & 0.1252 & 0.6884 & 0.0799 \\ 0.3408 & 0.0421 & 0.0667 & 0.0544 & 0.4959 \end{bmatrix}.$$

The matrix B is of a form in which we usually would get these matrices from applications. The left singular vectors of B are given by

$$\begin{bmatrix} -0.3558 & 0.6822 & 0.0367 & -0.6258 & -0.1220 \\ -0.5048 & -0.2770 & 0.0711 & -0.1661 & 0.7973 \\ -0.4216 & -0.2631 & 0.7608 & 0.0775 & -0.4100 \\ -0.5738 & -0.3030 & -0.6440 & 0.0368 & -0.4035 \\ -0.3339 & 0.5448 & -0.0006 & 0.7572 & 0.1357 \end{bmatrix}.$$

We now sort the second singular vector and obtain the permutation $(4, 2, 3, 5, 1)$*, which corresponds to the permutation matrix*

$$\tilde{P} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

We apply the permutation to the matrix B *and obtain*

$$\tilde{B} = \tilde{P}B\tilde{P}^T = \left[\begin{array}{ccc|cc} 0.6884 & 0.0344 & 0.1252 & 0.0799 & 0.0720 \\ 0.5436 & 0.1888 & 0.1788 & 0.0636 & 0.0252 \\ 0.3875 & 0.3154 & 0.2527 & 0.0015 & 0.0428 \\ \hline 0.0544 & 0.0421 & 0.0667 & 0.4959 & 0.3408 \\ 0.0192 & 0.0393 & 0.0887 & 0.6872 & 0.1655 \end{array}\right],$$

which exhibits the original block structure of $\tilde{A}$*.*

Once we know how to permute one large block that contains at least two subblocks into two blocks, we can employ this procedure recursively. We compute the second left singular vector, sort it and apply the permutation. Then, we check if the norm ($\mathbb{1}$-norm, see Definition 4.4) of the potential blocks is above a given threshold, see Definition 4.7. In this case, we have found two blocks and proceed recursively with each of the two blocks. Note that since we measure the size of the entries in the diagonal blocks and not in the off-diagonal blocks, the threshold can stay the same for all recursive calls. If the norm of the potential blocks is not above a given threshold, we cannot split the block any further and stop.

Algorithm 1: Identification of nearly decoupled blocks

Input: Matrix B, threshold $thr(= 1 - \delta)$
Output: Number m and sizes $n_i,\ i = 1, \ldots, m$ of identified blocks in B, a permutation matrix P such that $PBP^T = A + E$.

1 Compute the second left singular vector u_2 of B.
2 Sort it and use the resulting permutation P to permute the matrix B.
3 Identify two potential blocks B_1 and B_2 by using the change in sign in u_2.
4 The size of the first block is the number of negative values in u_2, the size of the second block the number of positive values in u_2.
5 **if** *the norm of the diagonal blocks is larger than thr* **then**
6 We have found two blocks and separate them.
7 Proceed recursively with step 1. applied to each of the blocks
8 **else**
9 The current block cannot be reduced any further.
10 Increase the counter of blocks by one. Stop.

6. Numerical tests. In this section, we present some numerical examples subdivided into three categories. First we discuss a constructed example, where we know the hidden structure and can see it fully recovered by our method. In the second subsection, we show results for a the molecule n-pentane, that was also used as example in [5], and in this case we obtain the same results. In the last subsection, we

present two slightly more challenging examples, where the algorithms in [5] and [6] have difficulties identifying the meta-stable states. With our algorithm, we are able to identify the maximal number of diagonal blocks such that the coupling matrix is diagonally dominant.

For numerical tests, Algorithm 1, was implemented in MATLAB® Version 7.0 and run on a PC with an Intel(R) Pentium(R) 4 CPU 3.20GHz processor. The relative machine precision was `eps` $= 2.2204 \times 10^{-16}$. In all example figures, we denote by n the number of unknowns and by nz the number of non-zero elements in the matrix.

6.1. A constructed example. The first example illustrates the ability of our method to recover a hidden block structure. It is constructed in the same manner as Example 5.1 in Section 5, is of size $n = 338$ and has $nz = 113906$ non-zero entries. In Figure 6.1, the upper left matrix is the original block diagonally dominant matrix, where we clearly can distinguish the diagonal blocks. The corresponding coupling matrix is diagonally dominant with values slightly larger than 0.5 on the diagonal. Hence, the order of the perturbation is 10^{-1}. The upper right matrix is a random symmetric permutation of the first matrix. Here, no structure can be seen. The lower left matrix depicts the recovered blocks after the calculation of one singular vector and the application of the corresponding permutation. One can see, that the block structure is to a large extent recovered but some parts are not fully restored yet. The lower right matrix now depicts the recovered structure after recursive application of the algorithm. We can see, that we have obtained the same blocks as in the original matrix up to permutation of the blocks and entries within a block. For some examples with such a large perturbation as in the previous example, the algorithm may fail as we can see in the next random example. Here, the structure can not be recovered as is depicted in Figure 6.2.

In general, one can say that the smaller the perturbation the better the algorithm recovers the hidden structure. We have tested 3 types of randomly generated examples. In the first type the diagonal entries of the coupling matrix are slightly larger than 0.5, in the second they are between 0.6 and 0.7, and in the third type the diagonal entries of the coupling matrix are about 0.9. We have run 1000 examples of each type. The structure could be recovered in the first case in 57,6%, in the second case in 85% and in the third case in 98,1% of all cases.

For comparison reasons, we have also run our algorithm using the right singular vectors instead of the left singular vectors. For the same 3 types of examples as in the previous paragraph, the optimal solution was found in the first case in 61,3%, in the second case in 84,9% and in the third case in 98,1% of all cases. However, as for some other problems, see, e.g., [13] and the references therein, the choice of left versus right singular vector might be important, see Sections 6.2 and 6.3.

The performance could be slightly enhanced by running the algorithm a second time using the right (left) singular vectors in case that it failed to find the optimal solution in the first run using left (right) singular vectors. In this case we obtained the optimal solution in the first case in 65,2%, in the second case in 87,8% and in the third case in 99,5% of all cases.

6.2. n-Pentane. The example of n-pentane was presented in [5, Section 5.2]. We will use this example to discuss the difference or the equivalence of the π-norm used in [5] and the $\mathbb{1}$-norm; see Section 4, Definition 4.4. This example is of size $n = 255$ and has $nz = 6464$ non-zero entries.

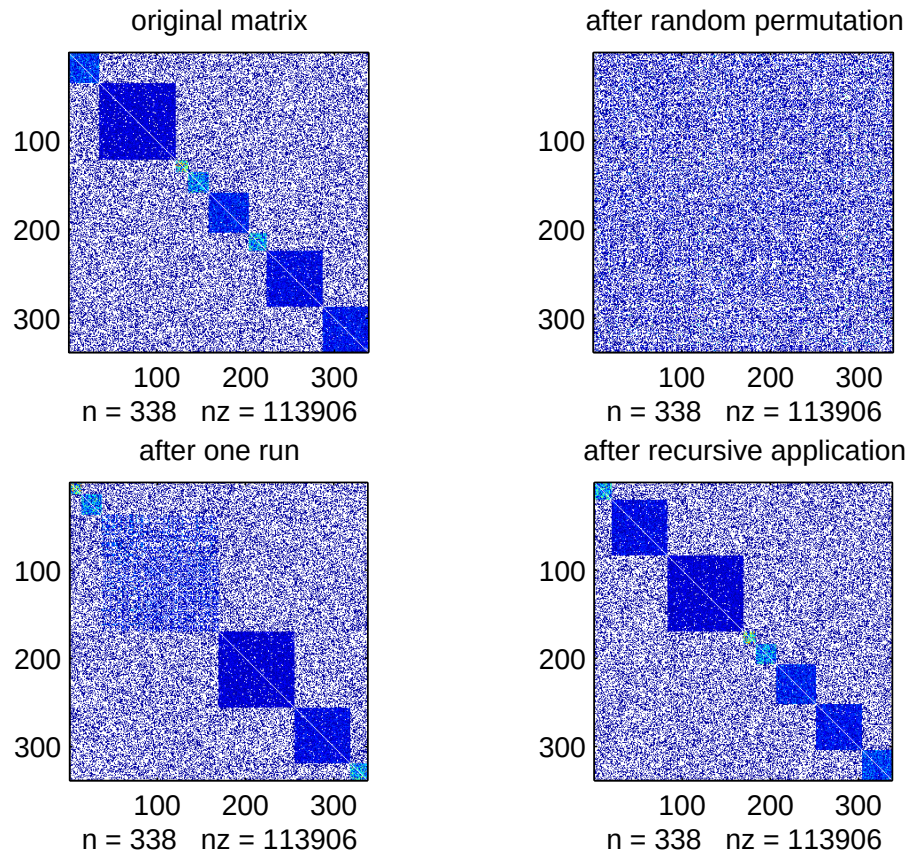

FIGURE 6.1. *Random example revealing 8 blocks*

In Figure 6.3 we illustrate the results of our algorithm using the $\mathbb{1}$-norm. We obtain 7 blocks of sizes 46, 24, 36, 20, 42, 47, 40 and the $\mathbb{1}$-norm coupling matrix, see Definition 4.3, is

$$W_{\mathbb{1}} = \begin{bmatrix} 0.6691 & 0.0278 & 0.1543 & 0.0018 & 0.0518 & 0.0517 & 0.1090 \\ 0.0061 & 0.6590 & 0.1543 & 0 & 0 & 0.0106 & 0.0462 \\ 0.0497 & 0.0809 & 0.8141 & 0 & 0.0009 & 0.0514 & 0.0003 \\ 0.0021 & 0 & 0 & 0.8178 & 0.0528 & 0.0283 & 0 \\ 0.0806 & 0 & 0.0019 & 0.0336 & 0.8478 & 0.0018 & 0.0465 \\ 0.1006 & 0.0031 & 0.0471 & 0.0885 & 0.0026 & 0.8027 & 0.0250 \\ 0.0350 & 0.1106 & 0.0084 & 0 & 0.0288 & 0.0092 & 0.8448 \end{bmatrix}.$$

We can see that the matrix is diagonally dominant. Now, we calculate the coupling matrix for our permuted matrix with the π-norm used in [5]. We again obtain a

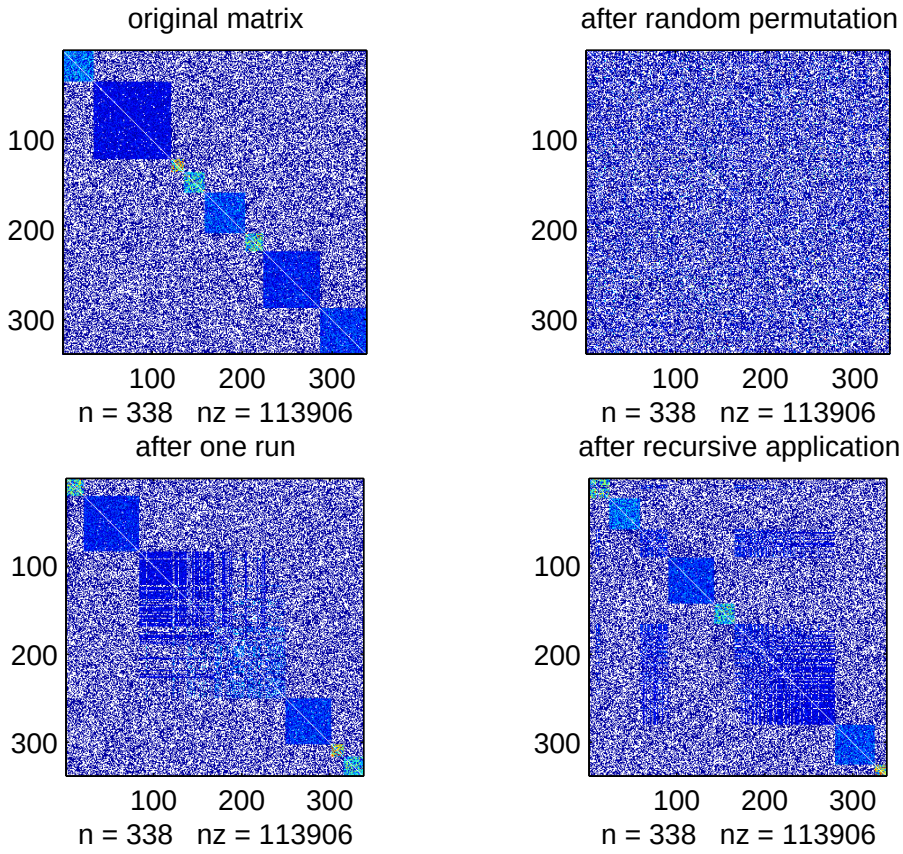

FIGURE 6.2. *Random example where the algorithm fails due to a very large error*

diagonally dominant matrix

$$W_\pi = \begin{bmatrix} 0.9777 & 0.0001 & 0.0053 & 0.0000 & 0.0044 & 0.0080 & 0.0046 \\ 0.0018 & 0.8763 & 0.0738 & 0 & 0 & 0.0007 & 0.0474 \\ 0.0190 & 0.0075 & 0.9727 & 0 & 0.0002 & 0.0005 & 0.0001 \\ 0.0008 & 0 & 0 & 0.9242 & 0.0365 & 0.0384 & 0 \\ 0.0169 & 0 & 0.0002 & 0.0033 & 0.9790 & 0.0002 & 0.0004 \\ 0.0311 & 0.0001 & 0.0006 & 0.0035 & 0.0002 & 0.9615 & 0.0031 \\ 0.0164 & 0.0048 & 0.0001 & 0 & 0.0004 & 0.0028 & 0.9755 \end{bmatrix},$$

yet, with larger values on the diagonal.

To compare the norms, we now run our algorithm using the π-norm. The result that we obtain is depicted in Figure 6.4.

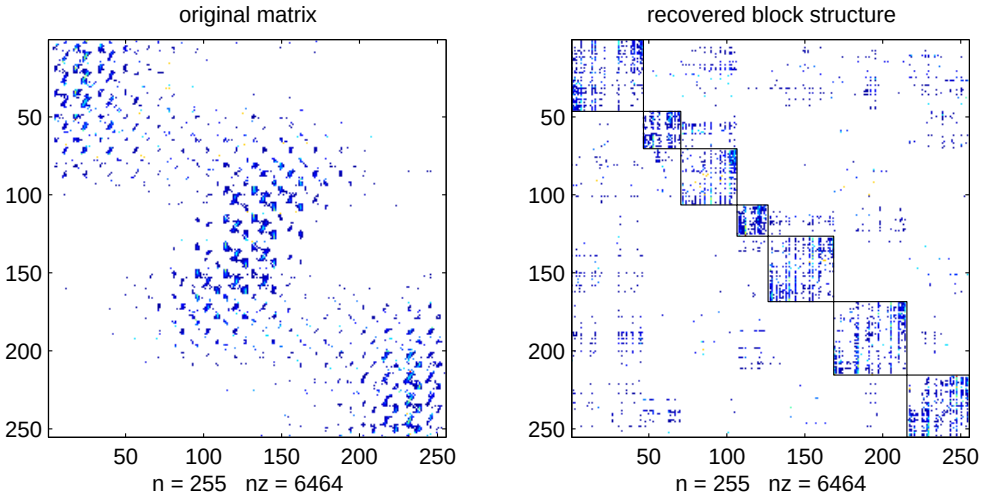

FIGURE 6.3. *n-Pentane (Ph300) revealing 7 blocks of sizes 46, 24, 36, 20, 42, 47, 40.*

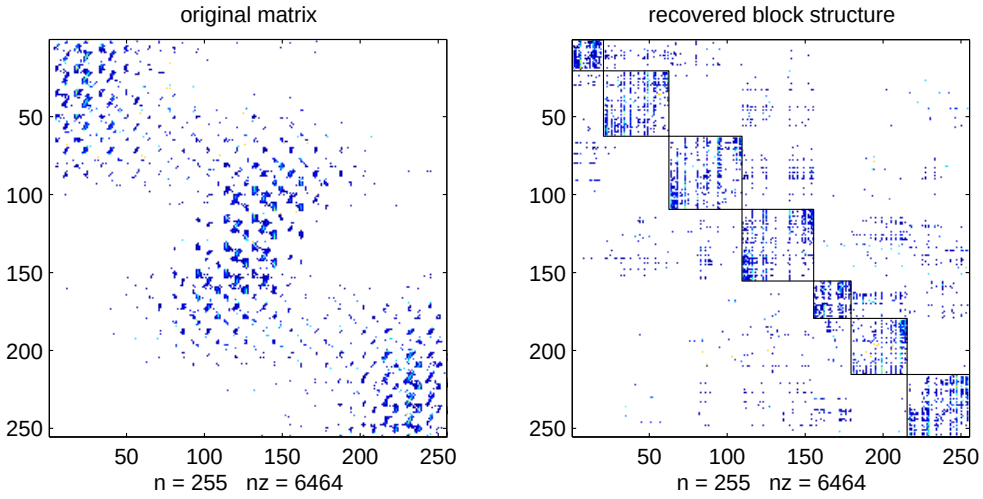

FIGURE 6.4. *n-Pentane (Ph300) revealing 7 blocks of sizes 20, 42, 47, 46, 24, 36, 40.*

The corresponding π-norm coupling matrix is

$$W_\pi = \begin{bmatrix} 0.9242 & 0.0365 & 0.0384 & 0.0008 & 0 & 0 & 0 \\ 0.0033 & 0.9790 & 0.0002 & 0.0169 & 0 & 0.0002 & 0.0004 \\ 0.0035 & 0.0002 & 0.9615 & 0.0311 & 0.0001 & 0.0006 & 0.0031 \\ 0.0000 & 0.0044 & 0.0080 & 0.9777 & 0.0001 & 0.0053 & 0.0046 \\ 0 & 0 & 0.0007 & 0.0018 & 0.8763 & 0.0738 & 0.0474 \\ 0 & 0.0002 & 0.0005 & 0.0190 & 0.0075 & 0.9727 & 0.0001 \\ 0 & 0.0004 & 0.0028 & 0.0164 & 0.0048 & 0.0001 & 0.9755 \end{bmatrix},$$

and the coupling matrix in the $\mathbb{1}$-norm is

$$W_{\mathbb{1}} = \begin{bmatrix} 0.8178 & 0.0528 & 0.0283 & 0.0021 & 0 & 0 & 0 \\ 0.0336 & 0.8478 & 0.0018 & 0.0806 & 0 & 0.0019 & 0.0465 \\ 0.0885 & 0.0026 & 0.8027 & 0.1006 & 0.0031 & 0.0471 & 0.0250 \\ 0.0018 & 0.0518 & 0.0517 & 0.6691 & 0.0278 & 0.1543 & 0.1090 \\ 0 & 0 & 0.0106 & 0.0061 & 0.6590 & 0.1543 & 0.0462 \\ 0 & 0.0009 & 0.0514 & 0.0497 & 0.0809 & 0.8141 & 0.0003 \\ 0 & 0.0288 & 0.0092 & 0.0350 & 0.1106 & 0.0084 & 0.8448 \end{bmatrix}.$$

We see that we obtain the same matrices up to permutation of the blocks. In this example it does not make any difference which norm we use for calculations.

Our next example is the same molecule but in a different temperature setting. This example is of size $n = 307$ and has $nz = 19116$ non-zero entries. Again, we first run the algorithm using the $\mathbb{1}$-norm. In this case, we obtain only 5 blocks depicted in Figure 6.5.

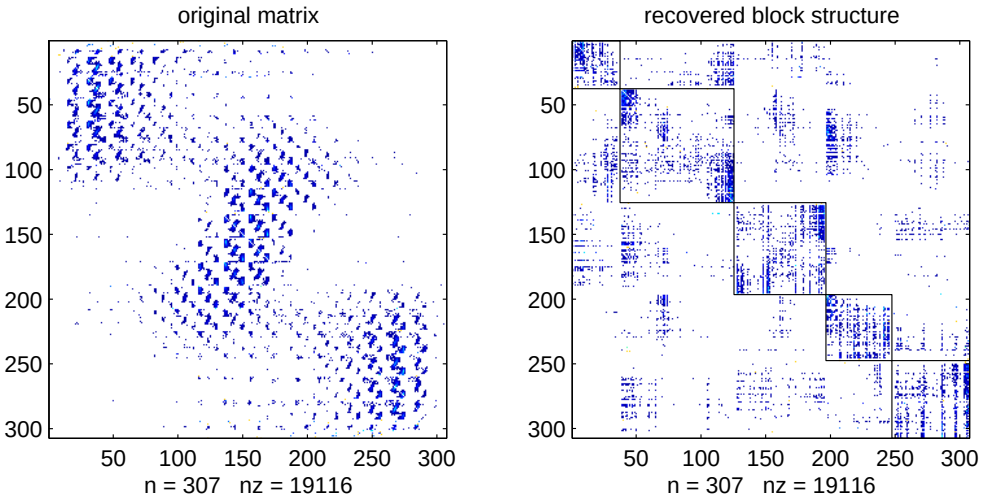

FIGURE 6.5. *n-Pentane (Ph500) revealing 5 blocks of sizes 37, 88, 71, 51, 60.*

The corresponding $\mathbb{1}$-norm coupling matrix is

$$W_{\mathbb{1}} = \begin{bmatrix} 0.5523 & 0.1207 & 0.0696 & 0.0166 & 0.0026 \\ 0.2283 & 0.6250 & 0.0884 & 0.2771 & 0.0691 \\ 0.1950 & 0.1476 & 0.6150 & 0.0401 & 0.0849 \\ 0.0047 & 0.0842 & 0.0250 & 0.6725 & 0.1224 \\ 0.0021 & 0.0894 & 0.1192 & 0.0842 & 0.6549 \end{bmatrix}.$$

It is diagonally dominant and the corresponding π-norm coupling matrix is

$$W_{\pi} = \begin{bmatrix} 0.5842 & 0.3024 & 0.1054 & 0.0058 & 0.0021 \\ 0.0333 & 0.8061 & 0.0682 & 0.0609 & 0.0314 \\ 0.0265 & 0.1559 & 0.7738 & 0.0274 & 0.0164 \\ 0.0023 & 0.2184 & 0.0430 & 0.6977 & 0.0386 \\ 0.0007 & 0.0933 & 0.0213 & 0.0320 & 0.8527 \end{bmatrix}.$$

We see that it also is diagonally dominant, again, with larger values on the diagonal.

For comparison, we now run the algorithm using the π-norm. We obtain 6 blocks, one more than with the $\mathbb{1}$-norm. The result is depicted in Figure 6.6.

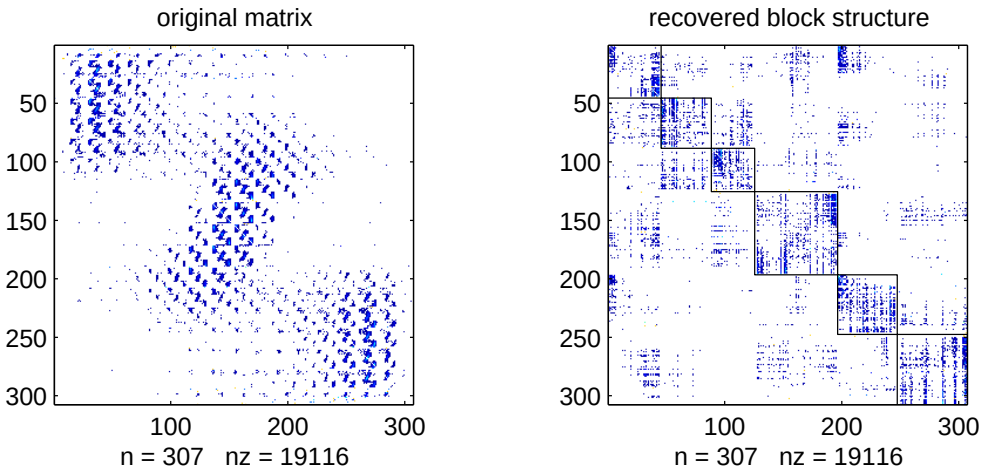

FIGURE 6.6. *n-Pentane (Ph500) revealing 6 blocks of sizes 45, 43, 37, 71, 51, 60.*

We see that we have the same blocks except that the block of size 88 is subdivided into two blocks of sizes 45 and 43. If we look at the coupling matrices, we can see the reason. The π-norm coupling matrix

$$W_\pi = \begin{bmatrix} 0.7642 & 0.0236 & 0.0036 & 0.0913 & 0.0808 & 0.0365 \\ 0.0628 & 0.7921 & 0.1123 & 0.0069 & 0.0079 & 0.0180 \\ 0.0240 & 0.2784 & 0.5842 & 0.1054 & 0.0058 & 0.0021 \\ 0.1516 & 0.0043 & 0.0265 & 0.7738 & 0.0274 & 0.0164 \\ 0.2107 & 0.0078 & 0.0023 & 0.0430 & 0.6977 & 0.0386 \\ 0.0787 & 0.0146 & 0.0007 & 0.0213 & 0.0320 & 0.8527 \end{bmatrix}$$

is diagonally dominant. Yet, the $\mathbb{1}$-norm coupling matrix

$$W_{\mathbb{1}} = \begin{bmatrix} 0.4897 & 0.0765 & 0.0354 & 0.1191 & 0.2204 & 0.0589 \\ 0.1576 & 0.5290 & 0.1593 & 0.0213 & 0.0980 & 0.0348 \\ 0.0365 & 0.2507 & 0.5523 & 0.1336 & 0.0228 & 0.0042 \\ 0.1585 & 0.0244 & 0.1016 & 0.6150 & 0.0288 & 0.0718 \\ 0.1329 & 0.0123 & 0.0034 & 0.0348 & 0.6725 & 0.1440 \\ 0.1105 & 0.0206 & 0.0013 & 0.1411 & 0.0716 & 0.6549 \end{bmatrix}$$

has a value smaller than 0.5 on the diagonal. Hence, we obtain the same qualitative results independently of the norm we use. The norm only affects the number of the determined blocks and, hence, the error tolerance.

For this example we also will show the result obtained by using right singular vectors, since it slightly improves the result obtained with left singular vectors. Here, we obtain 6 blocks as depicted in Figure 6.7 with both coupling matrices $W_{\mathbb{1}}$ and W_π

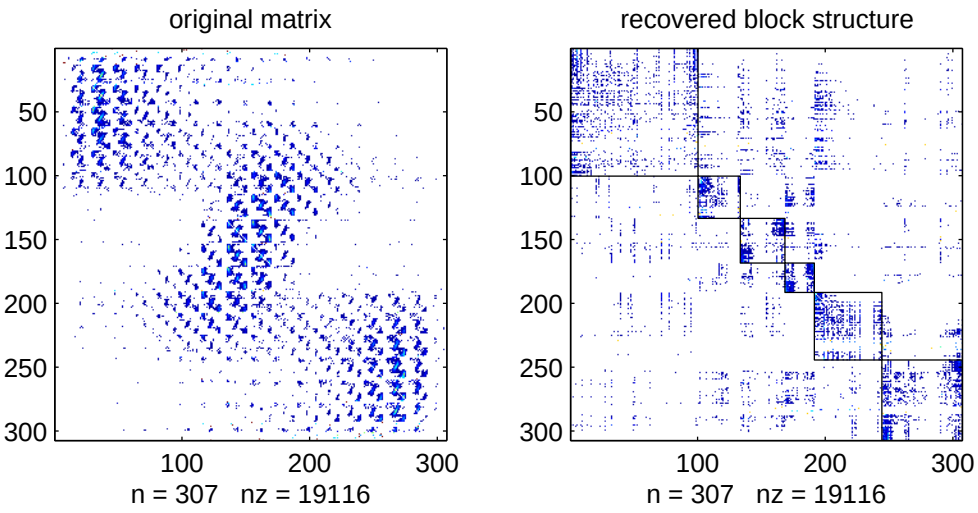

FIGURE 6.7. *Algorithm run using right singular vectors for n-Pentane (Ph500) revealing 6 blocks of sizes 100, 33, 35, 23, 53, 63.*

being diagonally dominant. The coupling matrices are given by

$$W_{\mathbb{1}} = \begin{bmatrix} 0.6227 & 0.0707 & 0.1015 & 0.0196 & 0.1288 & 0.0568 \\ 0.1454 & 0.5793 & 0.0233 & 0.2065 & 0.0009 & 0.0445 \\ 0.1636 & 0.0172 & 0.5072 & 0.1274 & 0.0935 & 0.0912 \\ 0.0323 & 0.0881 & 0.1341 & 0.6769 & 0.0042 & 0.0643 \\ 0.1019 & 0.0008 & 0.0509 & 0.0019 & 0.6597 & 0.1848 \\ 0.0598 & 0.0361 & 0.0967 & 0.0992 & 0.0776 & 0.6305 \end{bmatrix},$$

and

$$W_{\pi} = \begin{bmatrix} 0.6606 & 0.0234 & 0.1409 & 0.0068 & 0.1024 & 0.0658 \\ 0.1279 & 0.6557 & 0.0177 & 0.1828 & 0.0006 & 0.0153 \\ 0.1194 & 0.0027 & 0.7371 & 0.0359 & 0.0296 & 0.0752 \\ 0.0119 & 0.0584 & 0.0740 & 0.8441 & 0.0011 & 0.0105 \\ 0.1742 & 0.0002 & 0.0595 & 0.0011 & 0.7248 & 0.0402 \\ 0.0856 & 0.0036 & 0.1155 & 0.0078 & 0.0308 & 0.7566 \end{bmatrix}.$$

6.3. Two more difficult cases. In this section we present two cases where the algorithms presented in [5] and [6] have difficulties identifying meta-stable conformations. For the first matrix, that is of size $n = 158$ and has $nz = 24806$ non-zero entries, this is due to the fact that the algorithms in [5] and [6] identify the number of blocks by looking at spectral gaps. In this example, the spectrum of the matrix does not have any gaps. Our algorithm, on the opposite, does not need to know the number of blocks in advance but it is calculated in the process. We obtain the result depicted in Figure 6.8.

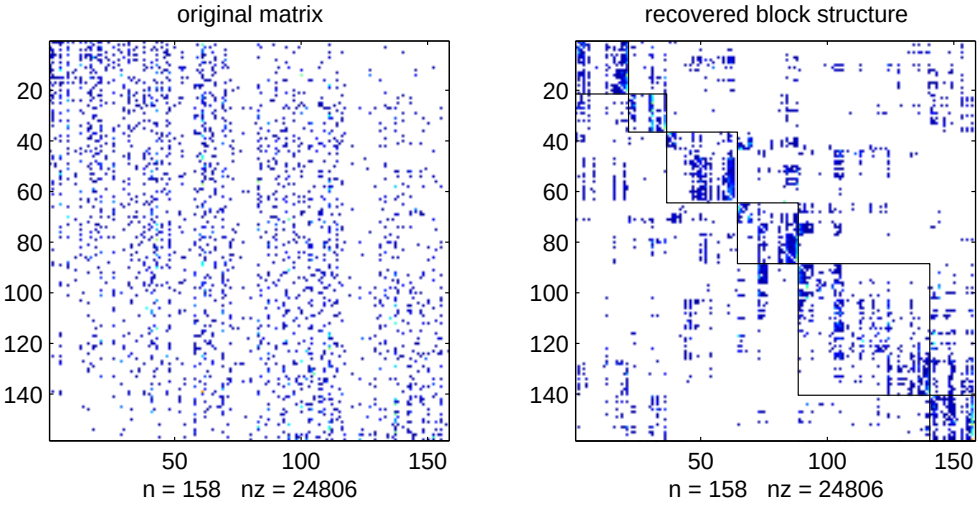

FIGURE 6.8. *A matrix without a spectral gap (Pmatrix2) revealing 6 blocks of sizes 28, 24, 21, 15, 52, 18.*

The corresponding $\mathbb{1}$-norm coupling matrix is

$$W_{\mathbb{1}} = \begin{bmatrix} 0.5823 & 0.1896 & 0.0852 & 0.0091 & 0.1324 & 0.0014 \\ 0.0892 & 0.6621 & 0.0139 & 0.0703 & 0.1463 & 0.0182 \\ 0.1115 & 0.0263 & 0.6413 & 0.0796 & 0.0944 & 0.0468 \\ 0.0133 & 0.0548 & 0.1161 & 0.6437 & 0.0349 & 0.1372 \\ 0.0507 & 0.0958 & 0.0491 & 0.0156 & 0.6369 & 0.1518 \\ 0.0007 & 0.0187 & 0.0477 & 0.1025 & 0.1364 & 0.6940 \end{bmatrix}.$$

The π-norm coupling matrix is

$$W_{\pi} = \begin{bmatrix} 0.6465 & 0.1410 & 0.0938 & 0.0066 & 0.1111 & 0.0010 \\ 0.0713 & 0.7484 & 0.0119 & 0.0346 & 0.1253 & 0.0086 \\ 0.0444 & 0.0111 & 0.8034 & 0.0489 & 0.0450 & 0.0473 \\ 0.0034 & 0.0356 & 0.0539 & 0.7828 & 0.0063 & 0.1180 \\ 0.0563 & 0.1256 & 0.0483 & 0.0062 & 0.6735 & 0.0901 \\ 0.0005 & 0.0078 & 0.0464 & 0.1052 & 0.0823 & 0.7578 \end{bmatrix}.$$

Now, again, we run the algorithm using the π-norm. We obtain two more blocks than with the $\mathbb{1}$-norm.

The π-norm coupling matrix is

$$W_{\pi} = \begin{bmatrix} 0.6465 & 0.1410 & 0.0938 & 0.0031 & 0.0036 & 0.0238 & 0.0872 & 0.0010 \\ 0.0713 & 0.7484 & 0.0119 & 0.0250 & 0.0096 & 0.0012 & 0.1240 & 0.0086 \\ 0.0444 & 0.0111 & 0.8034 & 0.0252 & 0.0237 & 0.0422 & 0.0028 & 0.0473 \\ 0.0032 & 0.0513 & 0.0553 & 0.5565 & 0.2221 & 0.0035 & 0.0037 & 0.1045 \\ 0.0037 & 0.0198 & 0.0524 & 0.2238 & 0.5632 & 0.0044 & 0.0011 & 0.1317 \\ 0.0333 & 0.0034 & 0.1248 & 0.0047 & 0.0059 & 0.5652 & 0.1221 & 0.1405 \\ 0.0694 & 0.1952 & 0.0047 & 0.0028 & 0.0008 & 0.0695 & 0.5962 & 0.0614 \\ 0.0005 & 0.0078 & 0.0464 & 0.0467 & 0.0585 & 0.0466 & 0.0358 & 0.7578 \end{bmatrix}.$$

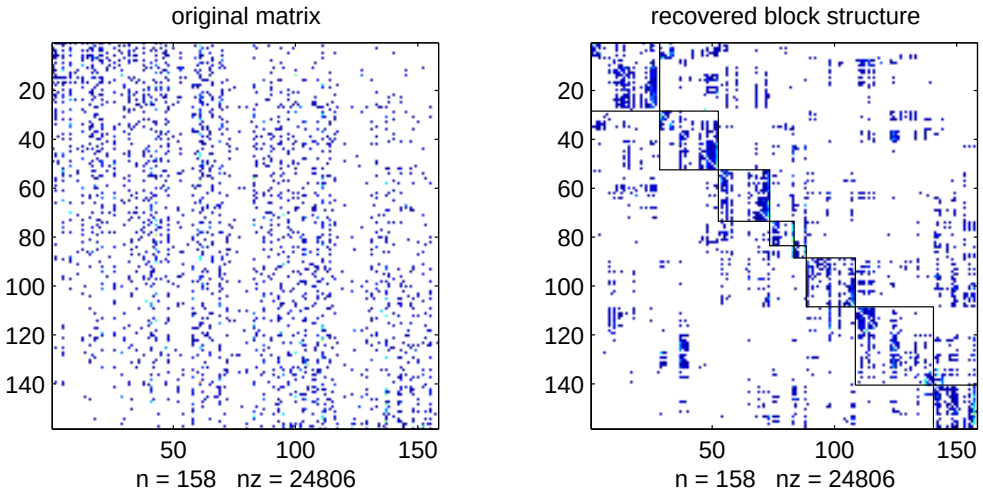

FIGURE 6.9. *A matrix without a spectral gap (Pmatrix2) revealing 8 blocks of sizes 28, 24, 21, 10, 5, 20, 32, 18.*

The corresponding $\mathbb{1}$-norm coupling matrix is given by

$$W_{\mathbb{1}} = \begin{bmatrix} 0.5823 & 0.1896 & 0.0852 & 0.0069 & 0.0022 & 0.0247 & 0.1076 & 0.0014 \\ 0.0892 & 0.6621 & 0.0139 & 0.0486 & 0.0217 & 0.0076 & 0.1387 & 0.0182 \\ 0.1115 & 0.0263 & 0.6413 & 0.0484 & 0.0312 & 0.0653 & 0.0291 & 0.0468 \\ 0.0167 & 0.0678 & 0.1302 & 0.5309 & 0.0762 & 0.0290 & 0.0187 & 0.1305 \\ 0.0066 & 0.0288 & 0.0879 & 0.2542 & 0.4627 & 0.0067 & 0.0026 & 0.1506 \\ 0.0331 & 0.0218 & 0.1155 & 0.0143 & 0.0083 & 0.4768 & 0.1369 & 0.1934 \\ 0.0618 & 0.1421 & 0.0075 & 0.0092 & 0.0021 & 0.0806 & 0.5708 & 0.1259 \\ 0.0007 & 0.0187 & 0.0477 & 0.0485 & 0.0540 & 0.0635 & 0.0729 & 0.6940 \end{bmatrix},$$

where we find exactly two values smaller than 0.5 on the diagonal.

Here, we also show a slightly better result obtained by using right singular vectors. Here, we obtain 8 blocks as depicted in Figure 6.10 with both coupling matrices $W_{\mathbb{1}}$ and W_{π} being diagonally dominant. The coupling matrices are given by

$$W_{\mathbb{1}} = \begin{bmatrix} 0.5823 & 0.1891 & 0.0743 & 0.0151 & 0.0486 & 0.0853 & 0.0043 & 0.0009 \\ 0.0871 & 0.6620 & 0.0108 & 0.0734 & 0.1395 & 0.0156 & 0.0026 & 0.0090 \\ 0.1031 & 0.0222 & 0.6761 & 0.0872 & 0.0259 & 0.0207 & 0.0605 & 0.0044 \\ 0.0162 & 0.0568 & 0.1151 & 0.6496 & 0.0685 & 0.0027 & 0.0355 & 0.0557 \\ 0.0303 & 0.1738 & 0.0194 & 0.0504 & 0.5201 & 0.0690 & 0.0576 & 0.0793 \\ 0.1578 & 0.0328 & 0.0274 & 0.0035 & 0.1463 & 0.5093 & 0.1153 & 0.0077 \\ 0.0135 & 0.0036 & 0.1503 & 0.0392 & 0.0767 & 0.0647 & 0.5027 & 0.1492 \\ 0.0013 & 0.0181 & 0.0068 & 0.1016 & 0.1721 & 0.0039 & 0.1312 & 0.5649 \end{bmatrix},$$

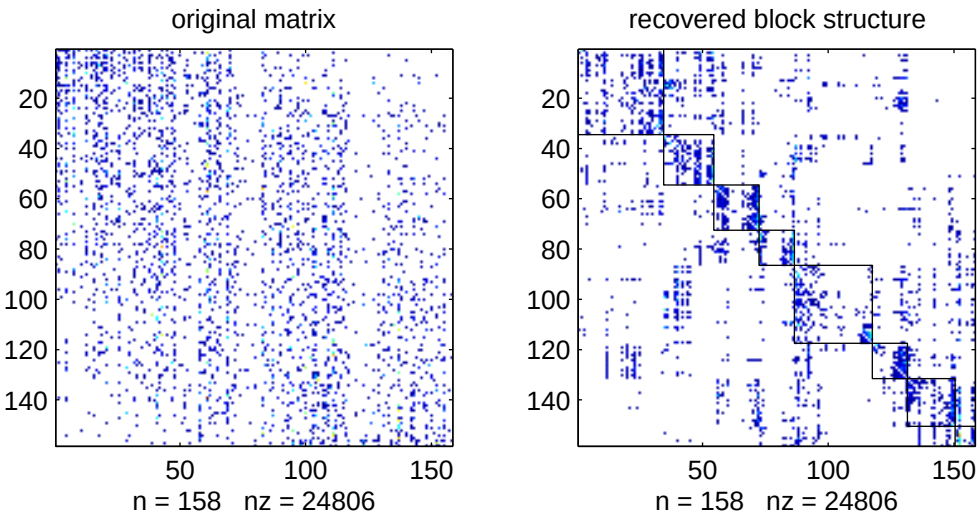

FIGURE 6.10. *Algorithm run using right singular vectors for a matrix without a spectral gap (Pmatrix2) revealing 8 blocks of sizes 34, 20, 18, 14, 31, 14, 19, 8.*

and

$$W_\pi = \begin{bmatrix} 0.6399 & 0.1453 & 0.0796 & 0.0071 & 0.0328 & 0.0911 & 0.0037 & 0.0005 \\ 0.0832 & 0.7368 & 0.0104 & 0.0347 & 0.1148 & 0.0155 & 0.0011 & 0.0035 \\ 0.0431 & 0.0098 & 0.8053 & 0.0492 & 0.0160 & 0.0044 & 0.0686 & 0.0036 \\ 0.0041 & 0.0354 & 0.0530 & 0.7824 & 0.0361 & 0.0003 & 0.0281 & 0.0607 \\ 0.0275 & 0.1682 & 0.0249 & 0.0519 & 0.5115 & 0.0669 & 0.0491 & 0.1001 \\ 0.1796 & 0.0534 & 0.0161 & 0.0011 & 0.1571 & 0.5097 & 0.0804 & 0.0025 \\ 0.0042 & 0.0022 & 0.1436 & 0.0546 & 0.0664 & 0.0463 & 0.5157 & 0.1671 \\ 0.0004 & 0.0057 & 0.0063 & 0.0977 & 0.1120 & 0.0012 & 0.1384 & 0.6384 \end{bmatrix}.$$

The second example that is of size $n = 1772$ and has $nz = 289432$ non-zero entries is problematic in a different way. Here, we have a very large cluster of eigenvalues very close to 1. This makes the matrix very badly conditioned, especially as far as the calculation of the stationary distribution is concerned. Also, the algorithms in [5] and [6] have difficulties identifying the right number of blocks. In Figure 6.11, we depict the results calculated with the $\mathbb{1}$-norm.

This is the maximum number of blocks that can be identified, such that the coupling matrix is diagonally dominant. Yet, blocks of very small sizes probably do not make a lot of sense from the chemical point of view. However, this is not a problem, since smaller blocks can always be merged into larger blocks. An alternative strategy could be to restrict the minimal block size in advance. In this case, we would only split up into blocks if they are of required size.

Since the coupling matrix is too large to be presented here, we only list the diagonal values in matrix form, to be read row-wise. The first matrix represents the

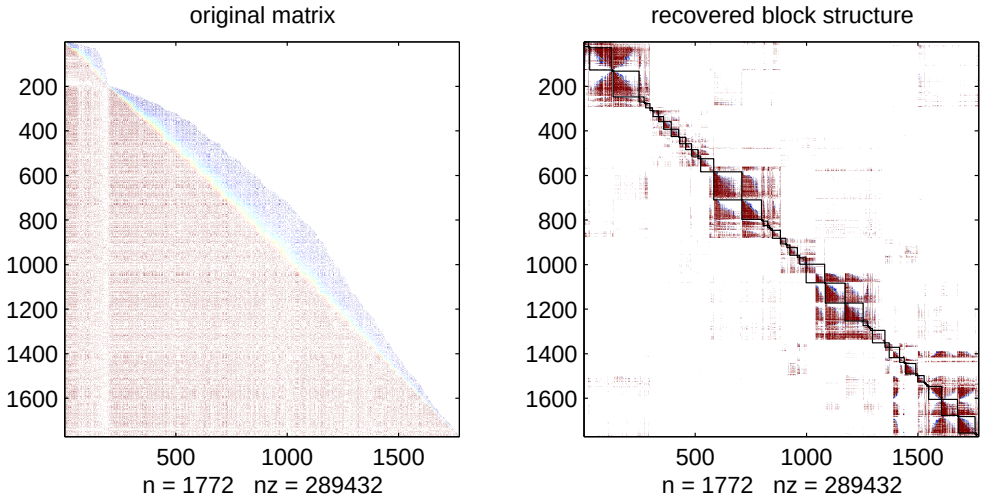

FIGURE 6.11. *A badly conditioned matrix with a large eigenvalue cluster very close to 1, revealing 65 blocks of sizes 2, 2, 2, 2, 2, 3, 3, 3, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5, 6, 6, 6, 7, 7, 8, 9, 9, 9, 11, 12, 12, 13, 19, 20, 20, 20, 21, 21, 23, 23, 25, 25, 26, 28, 29, 29, 29, 33, 34, 34, 36, 46, 46, 56, 58, 59, 72, 74, 81, 83, 88, 89, 100, 116, 126 sorted in ascending order.*

diagonal entries of the $W_{\mathbb{1}}$ coupling matrix

$$\begin{bmatrix} 0.7321 & 0.7062 & 0.7352 & 0.7638 & 0.8045 & 0.6664 & 0.6626 & 0.8248 \\ 0.6856 & 0.9896 & 0.9840 & 0.9203 & 0.9333 & 0.9529 & 0.9432 & 0.9292 \\ 0.9376 & 0.8892 & 0.9025 & 0.9364 & 0.8224 & 0.7813 & 0.7658 & 0.6851 \\ 0.8295 & 0.5526 & 0.7188 & 0.7615 & 0.6386 & 0.7036 & 0.9307 & 0.9909 \\ 0.9326 & 0.9197 & 0.9291 & 0.8951 & 0.7879 & 0.8306 & 0.7902 & 0.7644 \\ 0.7147 & 0.7073 & 0.6533 & 0.7263 & 0.7012 & 0.5946 & 0.9004 & 0.7416 \\ 0.8077 & 0.9154 & 0.9193 & 0.9036 & 0.7050 & 0.7008 & 0.7753 & 0.6902 \\ 0.5387 & 0.6347 & 0.7372 & 0.6714 & 0.7143 & 0.6478 & 0.7293 & 0.6636 \\ 0.6994 \end{bmatrix}.$$

The corresponding π-norm matrix has the following diagonal values

$$\begin{bmatrix} 0.9368 & 0.9311 & 0.9563 & 0.9695 & 0.9820 & 0.9876 & 0.7000 & 0.9867 \\ 0.9927 & 1.0000 & 0.9845 & 0.9723 & 0.9812 & 0.9913 & 0.9945 & 0.9929 \\ 0.9887 & 0.9939 & 0.9874 & 0.9888 & 0.9940 & 0.9633 & 0.9591 & 0.9721 \\ 0.9807 & 0.9770 & 0.9823 & 0.9863 & 0.9904 & 0.9934 & 0.9920 & 0.9961 \\ 0.9741 & 0.9878 & 0.9886 & 0.9842 & 0.9958 & 0.9531 & 0.9579 & 0.9597 \\ 0.9813 & 0.9863 & 0.7004 & 0.9939 & 0.9964 & 1.0000 & 0.9660 & 0.9827 \\ 0.9797 & 0.9851 & 0.9903 & 0.9885 & 0.9443 & 0.9462 & 0.9317 & 0.9941 \\ 0.2636 & 0.9847 & 0.9785 & 0.9784 & 0.9652 & 0.9657 & 0.9659 & 0.9887 \\ 0.9919 \end{bmatrix}.$$

The $W_{\mathbb{1}}$ matrix is diagonally dominant. However, the W_{π} matrix is not. The diagonal entry that is second last in the first column is smaller than 0.5. It corresponds to a block of size two that can be eliminated by merging two blocks or restricting the minimal block size.

For completeness, we also state in Figure 6.12 the results calculated with the π-norm. Here we obtain very similar blocks as with the $\mathbb{1}$-norm with some differences

in the smaller blocks. These differences are due to the bad condition of the matrix and hence, the stationary distribution. Yet, the bigger picture that we obtain is qualitatively the same.

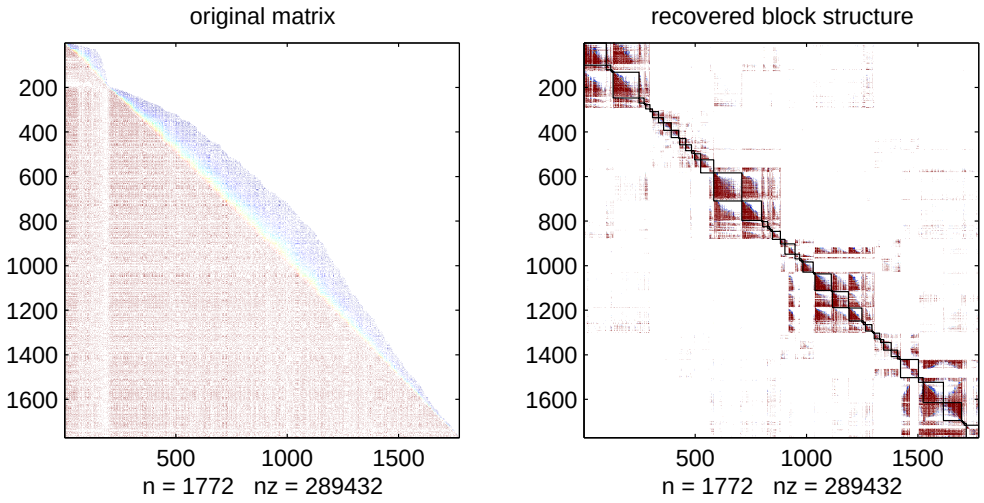

FIGURE 6.12. *A badly conditioned matrix with a large eigenvalue cluster very close to 1, revealing 64 blocks of sizes 2, 2, 2, 2, 2, 3, 3, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5, 6, 6, 6, 7, 7, 8, 9, 9, 9, 11, 12, 12, 13, 19, 20, 20, 21, 21, 23, 23, 23, 25, 26, 27, 27, 28, 29, 29, 33, 34, 34, 36, 46, 46, 56, 58, 59, 72, 74, 81, 81, 88, 91, 100, 116, 126 sorted in ascending order.*

The diagonal values of the W_π matrix are

$$\begin{bmatrix} 0.9563 & 0.9368 & 0.9311 & 0.9695 & 0.9820 & 0.9876 & 0.8552 & 0.9867 \\ 0.9927 & 1.0000 & 0.9850 & 0.9812 & 0.9913 & 0.9945 & 0.9929 & 0.9887 \\ 0.9939 & 0.9888 & 0.9872 & 0.9850 & 0.9940 & 0.9633 & 0.9591 & 0.9721 \\ 0.9822 & 0.9823 & 0.9863 & 0.9904 & 0.9934 & 0.9660 & 0.9827 & 0.9797 \\ 0.9885 & 0.9851 & 0.9903 & 0.6813 & 0.9847 & 0.9659 & 0.9657 & 0.9652 \\ 0.9785 & 0.9784 & 0.9887 & 0.9919 & 0.9325 & 0.9462 & 0.9317 & 0.9941 \\ 0.9920 & 0.9959 & 0.9741 & 0.9878 & 0.9842 & 0.9886 & 0.9905 & 0.9923 \\ 0.9813 & 0.9629 & 0.9597 & 0.9863 & 0.6515 & 0.9939 & 0.9964 & 1.0000 \end{bmatrix}.$$

The corresponding $W_{\mathbb{1}}$ matrix has the diagonal values

$$\begin{bmatrix} 0.7352 & 0.7321 & 0.7062 & 0.7638 & 0.8045 & 0.6664 & 0.6626 & 0.8248 \\ 0.6856 & 0.9896 & 0.9316 & 0.9333 & 0.9529 & 0.9432 & 0.9292 & 0.9376 \\ 0.8892 & 0.9364 & 0.9523 & 0.8946 & 0.8224 & 0.7813 & 0.7658 & 0.6851 \\ 0.5939 & 0.7188 & 0.7615 & 0.6386 & 0.7036 & 0.9004 & 0.7416 & 0.8077 \\ 0.9036 & 0.9154 & 0.9193 & 0.5387 & 0.6347 & 0.7293 & 0.6478 & 0.7143 \\ 0.7372 & 0.6714 & 0.6636 & 0.6994 & 0.7050 & 0.7008 & 0.7753 & 0.6902 \\ 0.9307 & 0.9909 & 0.9326 & 0.9197 & 0.8951 & 0.9291 & 0.9376 & 0.7819 \\ 0.7147 & 0.7960 & 0.7644 & 0.7073 & 0.6533 & 0.7263 & 0.7012 & 0.5946 \end{bmatrix}.$$

Here, again, both coupling matrices are diagonally dominant.

From the examples presented in this section we conclude that if the perturbation is not too large, a block diagonally dominant structure can be recovered or identified

by the proposed algorithm. Both norms, the $\mathbb{1}$-norm and the π-norm can be used for calculations. Qualitatively, one obtains similar results. Since for the real examples we do not know the "correct" answer we cannot decide, which norm is the "better" one. However, it is much cheaper to use the $\mathbb{1}$-norm and for a large eigenvalue cluster around one the calculation of the stationary distribution may be badly conditioned. Therefore, we suggest to use the $\mathbb{1}$-norm.

7. Conclusions. In this paper, we have presented a bisectioning algorithm for identifying meta-stable states of a Markov chain based on the calculation and sorting of the singular vector corresponding to the second largest singular value. The algorithm determines the maximum number of blocks such that the coupling matrix is diagonally dominant. One advantage of our approach is that it does not depend on a significant spectral gap in the transition matrix of the Markov chain. Thus, also matrices without a spectral gap or with a very large Perron cluster can be treated. Another advantage is that we calculate only two singular vectors instead of many eigenvectors. This allows to use iterative procedures such as Lanczos or Arnoldi iteration. Also, we suggest to abstain from using the stationary distribution since its calculation may be costly and badly conditioned. We show, that the same qualitative results can be achieved. This is demonstrated by numerical experiments.

Acknowledgements. The authors thank I. Slapnicar for the very interesting observation of sorting a singular vector. We are grateful to R. Nabben for fruitful discussions and for pointing out reference [10]. Furthermore, the authors thank A. Fischer and S. Kube for providing the example matrices.
This work is supported in part by DFG Research Center MATHEON in Berlin, by the U.S. National Science Foundation under grant CCF-0514489, and by the U.S. Department of Energy under grant DE-FG02-05ER25672.

REFERENCES

[1] C. J. Alpert, A. B. Kahng, and S.-Z. Yao. Spectral partitioning with multiple eigenvectors. *Discrete Appl. Math.*, 90(1-3):3–26, 1999.

[2] J. E. Atkins, E. G. Boman, and B. Hendrickson. A spectral algorithm for seriation and the consecutive ones problem. *SIAM J. Comput.*, 28(1):297–310 (electronic), 1999.

[3] Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, and H. van der Vorst, editors. *Templates for the solution of algebraic eigenvalue problems: a practical guide*. Software, Environments, and Tools. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000.

[4] A. Berman and R. J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994.

[5] P. Deuflhard, W. Huisinga, A. Fischer, and C. Schütte. Identification of almost invariant aggregates in reversible nearly uncoupled Markov chains. *Linear Algebra Appl.*, 315(1-3):39–59, 2000.

[6] P. Deuflhard and M. Weber. Robust Perron cluster analysis in conformation dynamics. *Linear Algebra Appl.*, 398:161–184, 2005.

[7] P. Drineas, A. Frieze, R. Kannan, S. Vempala, and V. Vinay. Clustering in large graphs and matrices. In *Proceedings of the Tenth Annual ACM-SIAM Symposium on Discrete Algorithms (Baltimore, MD, 1999)*, pages 291–299, New York, 1999. ACM.

[8] L. Elsner, C. R. Johnson, and M. Neumann. On the effect of the perturbation of a nonnegative matrix on its Perron eigenvector. *Czechoslovak Math. J.*, 32(107)(1):99–109, 1982.

[9] M. Fiedler. A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory. *Czechoslovak Math. J.*, 25(100)(4):619–633, 1975.

[10] S. Friedland and R. Nabben. On the second real eigenvalue of nonnegative and Z-matrices. *Linear Algebra Appl.*, 255:303–313, 1997.

[11] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, MD, 1983.

[12] T. Kato. *Perturbation theory for linear operators*. Classics in Mathematics. Springer-Verlag, Berlin, 1995. Reprint of the 1980 edition.

[13] Z. Kuang. The norm dependence of singular vectors. *J. Atmospheric Sci.*, 61(23):2943–2949, 2004.

[14] A. Pothen, H. D. Simon, and K.-P. Liou. Partitioning sparse matrices with eigenvectors of graphs. *SIAM J. Matrix Anal. Appl.*, 11(3):430–452, 1990.

[15] C. Schütte. http://biocomputing.mi.fu-berlin.de/.

[16] I. Slapnicar. Personal communication. 2004.

[17] D. A. Spielman and S.-H. Teng. Spectral partitioning works: planar graphs and finite element meshes. In *37th Annual Symposium on Foundations of Computer Science (Burlington, VT, 1996)*, pages 96–105. IEEE Comput. Soc. Press, Los Alamitos, CA, 1996.

[18] G. W. Stewart. On the sensitivity of nearly uncoupled Markov chains. In *Numerical solution of Markov chains*, volume 8 of *Probab. Pure Appl.*, pages 105–119. Dekker, New York, 1991.

[19] J. Xue. Blockwise perturbation theory for nearly uncoupled Markov chains and its application. *Linear Algebra Appl.*, 326(1-3):173–191, 2001.

[20] D. Zeimpekis and E. Gallopoulos. TMG: A MATLAB toolbox for generating term-document matrices from text collections. In J. Kogan, C. Nicholas, and M. Teboulle, editors, *Grouping Multidimensional Data: Recent Advances in Clustering*, pages 187–210. Springer, 2006.