

M. WEBER[†], T. GALLIAT[‡]

Characterization of Transition States in Conformational Dynamics using Fuzzy Sets

[†]Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB), Germany.
Internet: <http://www.zib.de/weber>

[‡]Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB), Germany.
Internet: <http://www.zib.de/galliat>

Characterization of Transition States in Conformational Dynamics using Fuzzy Sets

M. Weber¹ and T. Galliat¹

¹ Konrad-Zuse-Zentrum Berlin, Takustr. 7, 14195 Berlin, Germany

Abstract

Recently, a novel approach for the analysis of molecular dynamics on the basis of a transfer operator has been introduced [6, 2]. Therein conformations are considered to be disjoint metastable clusters within position space of a molecule. These clusters are defined by almost invariant characteristic functions that can be computed via *Perron Cluster* analysis [1].

The present paper suggests to replace crisp clusters with *fuzzy* clusters, i.e. to replace characteristic functions with membership functions. This allows a more sufficient characterization of transition states between different conformations and therefore leads to a better understanding of molecular dynamics. Furthermore, an indicator for the uniqueness of metastable fuzzy clusters and a fast algorithm for the computation of these clusters are described. Numerical examples are included.

Keywords. biochemical conformations, conformational dynamics, molecular dynamics, cluster analysis, transition states, fuzzy sets.

1 Introduction

The identification of metastable conformations on the basis of a transfer operator is an important concept to characterize the function of a molecule [6]. Using a suitable operator, these conformations correspond to metastable clusters within the position space of the molecule [2]. Their number and characteristic functions can be computed via Perron Cluster analysis [1].

Usually a molecule does not move directly between states of different conformations: There exist transition states which belong to different conformational changes and which have different frequencies of occurrence. Unfortunately the use of characteristic functions leads to crisp clusters, i.e. each transition state is assigned to exactly one cluster — a characterization that may be too strict. The aim of this paper is a replacement of characteristic functions by membership functions so that transition states can be assigned to different conformations with a certain degree of membership. It will be shown that the use of membership rules allows to construct an indicator for the uniqueness of the corresponding metastable clusters. Such an indicator is very useful in situations where Perron Cluster analysis leads to different possible sets of clusters.

In section 2 we will describe a concept of conformational analysis on the basis of a transfer operator approach. We will show that this concept leads to a special cluster problem.

In section 3 we are going to generalize this cluster problem by using membership functions instead of characteristic functions and in section 4 we will present a necessary and sufficient condition for its uniqueness.

In section 5 we will present a fast algorithm for the computation of unique metastable clusters based on membership functions. In addition, a simple indicator for uniqueness is given.

Finally, in section 6 two numerical examples are described.

2 Conformational analysis

Transfer operator approach. In classical molecular dynamics [5] a molecule with N atoms is represented by $d = 3N$ spatial coordinates $q \in \Omega \subset \mathbb{R}^d$ and $d = 3N$ momentum coordinates $p \in \mathbb{R}^d$. Molecular motion with respect to the time scale τ is modeled by the solution of a Hamiltonian differential equation, i.e. by the *flow* $\Phi^\tau : \Omega \times \mathbb{R}^d \rightarrow \Omega \times \mathbb{R}^d$ with

$$\Phi^\tau(q_0, p_0) = (q(\tau), p(\tau)).$$

Schütte et al. define the *spatial transition operator* $\mathbf{T} : L_{\mathcal{Q}}^n(\Omega) \rightarrow L_{\mathcal{Q}}^n(\Omega)$ as follows [2, 9]:

$$\mathbf{T}u(q) = \int_{\mathbb{R}^d} u(\pi_1(\Phi^{-\tau}(q, p))) \mathcal{P}(p) dp, \quad (1)$$

where

$$L_{\mathcal{Q}}^n(\Omega) := \{u : \Omega \rightarrow \mathcal{C}, \int_{\Omega} |u(q)|^n \mathcal{Q}(q) dq < \infty\}, \quad n = 1, 2,$$

$\pi_1 : \Omega \times \mathbb{R}^d \rightarrow \Omega$ is the projection $\pi_1(q, p) = q$ into spatial coordinates, \mathcal{P} is the normalized canonical momentum density and \mathcal{Q} is the normalized canonical spatial density.

This operator \mathbf{T} describes the momentum weighted fluctuations inside the canonical ensemble with respect to time scale τ . If there exists a non-void subset $A \subseteq \Omega$ with characteristic function $\chi_A : \Omega \rightarrow \{0, 1\}$ and

$$\mathbf{T}\chi_A = \chi_A, \quad (2)$$

then A is a *stable conformation* with regard to the flow Φ^τ . In coupled cases only χ_Ω meets (2). Therefore one is interested in *metastable conformations* $B \subset \Omega$ with almost \mathbf{T} -invariant characteristic functions χ_B , i.e. in characteristic functions that solve (2) approximately.

Galerkin discretization. For a numerical evaluation, one has to discretize the operator \mathbf{T} , i.e. one has to compute a decomposition of Ω into m disjoint subsets A_1, \dots, A_m and to replace Ω with a finite set $\Omega^* := \{A_1, \dots, A_m\}$ as described in [2, 8].

Hence one yields the positive stochastic (m, m) -matrix T

$$T_{i,j} := \frac{\int_{\Omega} \chi_{A_i}(q) \mathbf{T}\chi_{A_j}(q) \mathcal{Q}(q) dq}{\int_{\Omega} \chi_{A_i}(q) \mathcal{Q}(q) dq}. \quad (3)$$

Since T may be interpreted as the transfer operator of an underlying Markov chain, the computation of (3) can be done by *Hybrid Monte Carlo* (HMC) methods [4].

Cluster problem Let $\chi : \Omega^* \rightarrow \{0, 1\}$ be any characteristic function that solves

$$T\chi = \chi \quad (4)$$

approximately. If we set

$$A := \bigcup_{\{i \in \{1, \dots, m\}; \chi(A_i)=1\}} A_i, \quad (5)$$

then χ_A is an approximate solution of (2). Therefore the problem of finding metastable conformations corresponds to the following cluster problem: Compute s characteristic functions $\chi_1, \dots, \chi_s : \Omega^* \rightarrow \{0, 1\}$ so that

1. $\forall_{i=1, \dots, m} \exists_{j \in \{1, \dots, s\}} \chi_j(A_i) = 1$,
2. $\forall_{i=1, \dots, m} \forall_{(j,k) \in \{1, \dots, s\}; j \neq k} \chi_j(A_i) = 1 \Rightarrow \chi_k(A_i) = 0$,
3. $\forall_{j=1, \dots, s} T\chi_j \approx \chi_j$.

We will refer to the sets B_1, \dots, B_s that correspond to the characteristic functions χ_1, \dots, χ_s according to (5) as metastable clusters.

Perron Cluster analysis Deufelhard et al. suggest a solution for the above cluster problem by Perron Cluster analysis [1]. Let $v^1, \dots, v^s \in \mathbb{R}^m$ denote the eigenvectors corresponding to the eigenvalues of the Perron Cluster of the matrix T , i.e. corresponding to eigenvalues $\lambda \approx 1$. Assume s non-void disjoint index subsets $I_j \subset \{1, \dots, m\}$ so that the s -tuples $\{(v_i^1, \dots, v_i^s), i \in I_j\}$ share the same sign structure for every $j = 1, \dots, s$, then $\chi_j : \Omega^* \rightarrow \{0, 1\}$ with

$$\chi_j(A_i) := \begin{cases} 1 & \text{if } i \in I_j \\ 0 & \text{else} \end{cases}$$

is an approximate solution of our cluster problem. Note that one has to determine exactly s different sign structures.

Transition states From perturbation analysis it follows that components v_i^j being almost zero (absolute value beneath a problem depending threshold) are not “significantly large” enough to be assigned to any sign structure for sure [1]. In this case the corresponding A_i is a transition state between different metastable clusters. The following simple example illustrates the occurrence of transition states.

Example 2.1 Assume $\Omega^* = \{A_1, A_2, A_3\}$ and

$$T = \begin{pmatrix} 1 - \epsilon & \epsilon & 0 \\ a & 1 - a - b & b \\ 0 & \epsilon & 1 - \epsilon \end{pmatrix} \quad (6)$$

with $a, b > 0, a + b \leq 1, a + b \approx 1$ and $\epsilon > 0$ small. The matrix T is positive stochastic and has two eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 1 - \epsilon$ near 1 in the Perron Cluster, while the third eigenvalue $\lambda_3 = 1 - \epsilon - a - b$ is bounded away from 1. Therefore we have $s = 2$. The corresponding right eigenvectors are $v^1 = (1, 1, 1)^t$ and $v^2 = (b, 0, -a)^t$.

In this case λ_2 controls the stability of A_1 and A_3 , and v^2 controls the clustering via sign structure. Obviously A_2 is a transition state, because via an investigation of the sign structure of the second eigenvector we can not decide, whether A_2 belongs to A_1 or to A_3 .

3 Membership functions

Almost T -invariant cluster problem. We can generalize our cluster problem from section 2: For a given data set Ω and a linear operator \mathbf{T} find s characteristic functions $\chi_1, \dots, \chi_s : \Omega \rightarrow \{0, 1\}$ that satisfy approximately (2) and meet the following *covering property*

$$\sum_{i=1}^s \chi_i = \mathbb{1}_\Omega, \quad (7)$$

where $\mathbb{1}_\Omega : \Omega \rightarrow \{1\}$ is a constant function.

The Perron eigenspace. Let v^1, \dots, v^s be the eigenvectors corresponding to the Perron cluster $\lambda_1, \dots, \lambda_s$ of \mathbf{T} . Assuming $\lambda_1, \dots, \lambda_s \approx 1$ every

$$\chi \in I_T := \text{span}\{v^1, \dots, v^s\}$$

solves (2) approximately. I_T is called the *Perron eigenspace* of \mathbf{T} .

If there are s characteristic functions $\chi_1, \dots, \chi_s \in I_T$ which meet the covering property (7), then a solution of our cluster problem is given by a simple linear transformation of the eigenvectors v^1, \dots, v^s .

In general there exists no such linear transformation, because the codomain of characteristic functions is discrete. But extending the codomain to the interval $[0, 1]$ makes such a linear transformation possible. This changes the point of view from *crisp clustering* into *fuzzy clustering*.

From now on

$$\chi : \Omega \rightarrow [0, 1]$$

are *membership functions* which are also called *fuzzy sets* in the literature [10]. The sought-after subset $\Gamma_T \subset I_T$ is called the set of *almost T -invariant membership functions*. It is the intersection of the set of membership functions with the vector space I_T .

Complete system of membership functions. We are interested in linear independent membership functions $\chi_1, \dots, \chi_s \in \Gamma_T$. We call such a system *complete* if it meets the covering property (7) and if it is a basis of I_T , i.e. if it is a solution of our almost T -invariant cluster problem. Note that for the existence of complete systems it is necessary that $\mathbb{1}_\Omega \in \Gamma_T$.

Suppose $\mathbb{1}_\Omega \in \Gamma_T$ and that the eigenvectors $v^1, \dots, v^s : \Omega \rightarrow \mathbb{R}$ spanning I_T are bounded. Then the following linear transformation (LT1) leads to a complete system of membership functions in Γ_T :

- **Step 1: Positiveness.** Assume without loss of generality $v^1 = \mathbb{1}_\Omega$. For $i = 2, \dots, s$ define

$$w^i := v^i - (\inf_{\Omega} v^i) \mathbb{1}_\Omega. \quad (8)$$

Hence for $i = 2, \dots, s$ we have $w^i \in I_T$ and $\inf_{\Omega} w^i = 0$.

- **Step 2: Scaling.** Determine

$$\mu := \left(\sup_{\Omega} \left(\sum_{i=2}^s w^i \right) \right)^{-1}. \quad (9)$$

Hence for $i = 2, \dots, s$ we have $\sup_{\Omega} \mu w^i \leq 1$.

- **Step 3: Completing.** For $i = 2, \dots, s$ set

$$\chi_i := \mu w^i \quad (10)$$

and

$$\chi_1 := \mathbb{1}_{\Omega} - \sum_{i=2}^s \chi_i. \quad (11)$$

Then $\chi_1, \dots, \chi_s \in \Gamma_T$ is a complete system of membership functions with $\inf_{\Omega} \chi_i = 0$ for $i = 1, \dots, s$.

Example 3.1 Remember example 2.1. We show how the above linear transformation (LT1) works for $v^1 = (1, 1, 1)^t$ and $v^2 = (b, 0, -a)^t$. Step 1 leads to $w^2 = (b + a, a, 0)^t$. Step 2 gives us $\mu = 1/(b + a)$. Finally we get in step 3 a complete system of membership functions in Γ_T :

$$\chi_1 = \begin{pmatrix} 0 \\ \frac{b}{a+b} \\ 1 \end{pmatrix}, \quad \chi_2 = \begin{pmatrix} 1 \\ \frac{a}{a+b} \\ 0 \end{pmatrix}. \quad (12)$$

Thus a reasonable grade of membership is calculated for A_2 with regard to χ_1 and χ_2 respectively.

The above example shows that almost T -invariant membership functions assign transition states to different clusters with different grades of membership.

Indecomposable membership functions Look at the following two complete systems of membership functions for $\Omega^* = \{A_1, A_2, A_3\}$:

$$\left\{ \chi_1^* = \begin{pmatrix} 0.56 \\ 0.6 \\ 0.4 \end{pmatrix}, \chi_2^* = \begin{pmatrix} 0.44 \\ 0.4 \\ 0.6 \end{pmatrix} \right\}, \quad \left\{ \chi_1^{**} = \begin{pmatrix} 0.8 \\ 1 \\ 0 \end{pmatrix}, \chi_2^{**} = \begin{pmatrix} 0.2 \\ 0 \\ 1 \end{pmatrix} \right\}. \quad (13)$$

Both systems span the same vector space I_T , but the first system is *decomposable* (e.g. $\chi_1^* = 0.6\chi_1^{**} + 0.4\chi_2^{**}$), whereas the second one is *indecomposable*. This motivates the following definition:

Definition 3.2 (Indecomposable Membership Functions (IMF))

$$\Gamma_T^{\circ} := \{ \chi \in \Gamma_T; \chi_A, \chi_B \in \Gamma_T, \chi_A + \chi_B = \chi \Rightarrow \chi_A, \chi_B \text{ linear dependent} \}$$

is the set of indecomposable membership functions with regard to $\Gamma_T \subset I_T$.

Obviously $\mathbf{0}_{\Omega} : \Omega \rightarrow \{0\}$ is an IMF. Fuzzy sets that represent a mixture of different clusters like (χ_1^*, χ_2^*) can be split into linear independent membership functions like $(\chi_1^{**}, \chi_2^{**})$. Therefore Definition 3.2 means that IMFs are not mixtures of different clusters, they are *pure*. To get more familiar with IMFs we present some of their properties.

Lemma 3.3 Let $\chi \in \Gamma_T^o$ an indecomposable membership function $\chi : \Omega \rightarrow [0, 1]$, then the following propositions hold:

- (i) $\mu \in \mathbb{R}^+, \mu\chi \in \Gamma_T \Rightarrow \mu\chi \in \Gamma_T^o$
- (ii) $\mathbb{1}_\Omega \in \Gamma_T^o \Rightarrow \Gamma_T = \{\mu\mathbb{1}_\Omega; \mu \in [0, 1]\}$
- (iii) χ and $\mathbb{1}_\Omega$ linear independent $\Rightarrow \inf_\Omega(\chi) = 0$.

In particular if $\dim I_T > 1$, then for $\chi \in \Gamma_T^o$ there exists $\mu \in \mathbb{R}$ so that $\mu\chi \in \Gamma_T^o$, $\inf_\Omega(\mu\chi) = 0$ and $\sup_\Omega(\mu\chi) = 1$ (maximum scalability).

Proof: (i). If there exists $\chi_A, \chi_B \in \Gamma_T$ with $\mu\chi = \chi_A + \chi_B$, linear independence follows from $\mu^{-1}\chi_A + \mu^{-1}\chi_B = \chi$ and $\chi \in \Gamma_T^o$.

(ii). Follows from the decomposition $\mathbb{1}_\Omega = (\mathbb{1}_\Omega - \chi) + \chi$ for $\chi \in \Gamma_T$.

(iii). If $\inf_\Omega(\chi) = \delta > 0$, then one yields the decomposition $\chi = (\delta\mathbb{1}_\Omega) + (\chi - \delta\mathbb{1}_\Omega)$ with $(\chi - \delta\mathbb{1}_\Omega) \in \Gamma_T$ and therefore $\chi \notin \Gamma_T^o$. \square

After these preparations we are ready to define, how solutions of our cluster problem should look like.

Definition 3.4 (Indecomposable solution of almost T -invariant cluster problems)

If a system $\{\chi_1, \dots, \chi_s\}$ of membership functions $\chi_i : \Omega \rightarrow [0, 1]$ is complete and $\chi_1, \dots, \chi_s \in \Gamma_T^o$, then it is called an indecomposable solution of the almost T -invariant cluster problem.

An indecomposable solution is called *unique* if there are no other indecomposable membership functions in Γ_T^o , except for multiples of χ_1, \dots, χ_s according to Lemma 3.3(i).

4 Uniqueness of indecomposable cluster solutions

Extreme Element. We first prove a sufficient condition (*extreme element condition*) for the uniqueness of an indecomposable solution of the almost T -invariant cluster problem.

Definition 4.1 (Extreme Element) Let $S = \{\chi_1, \dots, \chi_s\}$ be any set of membership functions $\chi_i : \Omega \rightarrow [0, 1]$, then $x \in \Omega$ is called an extreme element of membership function χ_i with regard to S if $\chi_i(x) = 1$ and $\chi_j(x) = 0, j \neq i$.

Theorem 4.2 Let $S = \{\chi_1, \dots, \chi_s\}$ be any basis of membership functions $\chi_i : \Omega \rightarrow [0, 1]$ of the vector space I_T . Furthermore let $\{x_1, \dots, x_s\} \subset \Omega$ and x_i any extreme element of χ_i with regard to S for every $i = 1, \dots, s$. Then S is the unique indecomposable cluster solution.

By the linear transformation (LT1) (page 4) and the fact that $\chi_i : \Omega \rightarrow [0, 1]$ ($i = 1, 2$) attains its minimum 0 at some x_i for finite Ω and $\chi_1 + \chi_2 = \mathbb{1}_\Omega$, a direct consequence of 4.2 is the following corollary:

Corollary 4.3 *Let Ω finite and $\dim I_T = 2$, then there exists an unique indecomposable cluster solution that can be computed via transformation (LT1).*

Proof of 4.2: Because $\mathbf{1}_\Omega \in I_T$ and S is a basis of I_T with extreme elements for every $\chi_i \in S$ and $\mathbf{1}_\Omega$ has a unique representation as a linear combination of χ_1, \dots, χ_s , S meets the covering property (7).

To show that each $\chi_i \in S$ is indecomposable we consider $\chi_1 \in S$ without loss of generality. Let $\chi_A, \chi_B \in \Gamma_T$ with $\chi_1 = \chi_A + \chi_B$. Because S is a basis of I_T and there are extreme elements for every χ_i we get real positive numbers $\alpha_1, \dots, \alpha_s \in \mathbb{R}_0^+$ and $\beta_1, \dots, \beta_s \in \mathbb{R}_0^+$ with

$$\chi_A = \sum_{i=1}^s \alpha_i \chi_i, \quad \chi_B = \sum_{i=1}^s \beta_i \chi_i.$$

We further get

$$\chi_1 = \sum_{i=1}^s (\alpha_i + \beta_i) \chi_i.$$

Since S is a basis of I_T , we have $(\alpha_1 + \beta_1) = 1$ and $(\alpha_j + \beta_j) = 0$ for $j = 2, \dots, s$, where α_j and β_j are positive. This yields

$$\chi_A = \alpha_1 \chi_1, \quad \chi_B = \beta_1 \chi_1$$

and therefore linear dependency of χ_A and χ_B .

To prove uniqueness, suppose there exists another indecomposable solution $\tilde{S} = \{\tilde{\chi}_1, \dots, \tilde{\chi}_s\}$. We only have to show that $\tilde{S} \subset S$. Let $\chi \in \tilde{S}$. Since S is a basis of I_T and there are extreme elements for every $\chi_i \in S$, we have positive real numbers $\alpha_1, \dots, \alpha_s \in \mathbb{R}_0^+$ with

$$\chi = \sum_{i=1}^s \alpha_i \chi_i.$$

Further there exists a $j \in \{1, \dots, s\}$ with $\alpha_j \neq 0$ yielding

$$\chi = \alpha_j \chi_j + \sum_{i \neq j} \alpha_i \chi_i.$$

Obviously, $\chi \in \Gamma_T^o$ guarantees linear dependency of

$$\chi_A := \alpha_j \chi_j \in \Gamma_T \quad \text{and} \quad \chi_B := \sum_{i \neq j} \alpha_i \chi_i \in \Gamma_T$$

implying $\chi = \alpha_j \chi_j$. Since \tilde{S} meets the covering property, we have $\alpha_j = 1$ and therefore $\chi \in S$. \square

Whereas an indecomposable solution for $s = 2$ can be found simply by applying (LT1), corollary 4.3 does not hold in general for $s \geq 3$. In this case, after (LT1) a further linear transformation (LT2) has to be applied to the complete system of membership functions:

If $\Omega = \{A_1, \dots, A_m\}$ is finite, then a complete system S of membership functions χ_1, \dots, χ_s can be written as (m,s)-matrix C :

$$C_{i,j} := \chi_j(A_i), \quad i = 1, \dots, m, \quad j = 1, \dots, s. \quad (14)$$

Because S is complete the columns of C are linear independent and for the k -vector $\mathbf{1}_k := (1, \dots, 1)^t$ we have

$$C\mathbf{1}_s = \mathbf{1}_m. \quad (15)$$

The linear transformation (LT2) is defined as a multiplication of C with a (s,s) -matrix D . Since after this transformation (LT2) CD should again represent a complete system of membership functions, D has to be regular and to satisfy:

$$CD\mathbf{1}_s = \mathbf{1}_m \Rightarrow D\mathbf{1}_s = \mathbf{1}_s \Rightarrow D^{-1}\mathbf{1}_s = \mathbf{1}_s. \quad (16)$$

Feasibility of (LT2) due to the conditions $1 \geq (CD)_{i,j} \geq 0$ and (16) can geometrically be interpreted: The rows of C as points in \mathbb{R}^s are convex combinations of the rows of D^{-1} as points in \mathbb{R}^s . Because of (15) and (16) the rows of C and D^{-1} (as points) lie on the same hyper-plane in \mathbb{R}^s and by omitting the first column (orthogonal projection) of C and D^{-1} they can be interpreted as points in $\mathbb{R}^{(s-1)}$. Name these sets of points \mathcal{C} for matrix C and \mathcal{D} for matrix D^{-1} respectively. Let V_s be the convex hull of \mathcal{D} , then V_s is a $(s-1)$ -simplex with nonempty interior, because D is regular. $\mathcal{C} = \{y_1, \dots, y_m\}$ is called an I_T -projection of Ω . \mathcal{C} is not unique, because the choice of a basis of I_T is not unique.

Therefore feasibility of (LT2) is the same as

$$\text{co}(\mathcal{C}) \subseteq V_s \quad (17)$$

with the convex hull $\text{co}(\mathcal{C})$ of \mathcal{C} .

Uniqueness. If $\text{co}(\mathcal{C})$ is a $(s-1)$ -simplex, then we may choose $V_s = \text{co}(\mathcal{C})$ to yield a feasible transformation (LT2), i.e. there exist s points in \mathcal{C} that are related to s elements $x_1, \dots, x_s \in \Omega$ which span V_s . After computing CD we get via reversal of (14) a complete system of membership functions $S = \{\chi_1, \dots, \chi_s\} \subset \Gamma_T$, where x_i is an extreme element of χ_i for $i = 1, \dots, s$. Theorem 4.2 shows that S is the unique indecomposable cluster solution in this case.

The next theorem shows the relation between $\text{co}(\mathcal{C})$ being a simplex and the uniqueness of an indecomposable solution of I_T .

Theorem 4.4 *Let Ω finite. For given s -dimensional I_T there exists an unique indecomposable solution of the T -invariant cluster problem if and only if the convex hull $\text{co}(\mathcal{C}) \subset \mathbb{R}^{(s-1)}$ of an I_T -projection of Ω is a $(s-1)$ -simplex.*

Proof: One direction of 4.4 has already been shown. Therefore, we only have to prove that for $\text{co}(\mathcal{C})$ not being a $(s-1)$ -simplex there are —except for scalar multiplicity— more than s indecomposable membership functions of I_T .

We show that each facet of $\text{co}(\mathcal{C})$ leads to an indecomposable membership function. A facet F of $\text{co}(\mathcal{C})$ contains of $(s-1)$ linear independent points $y_1, \dots, y_{s-1} \in \mathcal{C}$ belonging to $(s-1)$ points $x_1, \dots, x_{s-1} \in \Omega$.

Take any outer simplex V_s with (17) and having F as a part of one of its facets. Apply the corresponding linear transformation (LT2). We yield s linear independent membership functions χ_1, \dots, χ_s . For one of them, say χ_s , we have

$$\chi_s(x_1), \dots, \chi_s(x_{s-1}) = 0. \quad (18)$$

Now suppose $\chi_s = \chi_A + \chi_B$ with

$$\chi_A = \sum_{j=1}^s \alpha_j \chi_j.$$

Hence for $i = 1, \dots, s-1$ with (18) and the positiveness of χ_s, χ_A and χ_B we have

$$0 = \sum_{j=1}^{s-1} \alpha_j \chi_j(x_i),$$

which is a regular system of $s-1$ linear equations. Therefore $\alpha_j = 0$ for $j = 1, \dots, s-1$ which implies indecomposability of χ_s . \square

Remark. (LT2) can be applied to the matrix of eigenvectors without applying (LT1) first. (17) converts the problem of finding a feasible linear transformation into the problem of finding an outer simplex V_s for a given set \mathcal{C} .

5 Inner Simplex Algorithm

Outer simplex algorithms. There are many different objective functions to convert the problem of finding a “best fitting” outer simplex for the set \mathcal{C} into a global optimization problem. But any of these problems will not be convex. Until now there is no possibility to solve any non-convex global optimization problem in reasonable time. This is a common property of the class of NP -hard problems [7]. However, for local optimization methods one may use the fact that after applying (LT1) a possible outer simplex for \mathcal{C} is spanned by the origin and the $(s-1)$ unit vectors in $\mathbb{R}^{(s-1)}$.

Inner simplex algorithm. Suppose there exists an unique indecomposable cluster solution. From Theorem 4.4 we know that we can compute this solution by defining D via the simplex $V_s := co(\mathcal{C})$ (see (16)). To determine V_s one may apply the following algorithm (*inner simplex algorithm*) based on ideas of the quickhull algorithm [3]:

1. Find two points $y_1, y_2 \in \mathcal{C}$ having maximum Euclidean distance with regard to every pair of points in \mathcal{C} . Obviously $y_1, y_2 \in co(\mathcal{C})$. Add y_1 and y_2 to the vertex set of the searched $(s-1)$ -simplex V_s .
2. For step counter $k = 2, \dots, (s-1)$: Find a point $x \in \mathcal{C}$ maximizing the Euclidean distance between point $(y - y_1)$ and the vector space spanned by $(y_2 - y_1), \dots, (y_k - y_1)$ via orthogonal projection. Obviously we have $y_{k+1} := y \in co(\mathcal{C})$. Therefore add y_{k+1} to the vertex set of the searched $(s-1)$ -simplex V_s .

After applying the inner simplex algorithm one yields V_s .

Transition states and intermediate states. To achieve an indicator for the fulfillment of the assumption that there exists a unique indecomposable solution, one applies (LT2) via inner simplex algorithm and looks at the lowest value θ of all membership functions. If and only if $\theta = 0$ this assumption was correct. $|\theta| \in [0, 1]$ is called an *indicator for uniqueness*.

If $\theta = 0$ one yields two classes of states in Ω after applying (LT2):

- *Extreme elements.* Look at Definition 4.1. Extreme elements are related to the vertices of V_s .
- *Transition states.* All elements in Ω that are not extreme elements are transition states. Their grade of membership to one of the extreme elements is defined via membership functions.

If $|\theta| > 0$ a third kind of states in Ω appears after applying (LT2). An element $x \in \Omega$ is called *intermediate state* if for one of the “membership functions” χ obtained after (LT2) with the inner simplex algorithm we have $\chi(x) < 0$.

These states are responsible for the ambiguity of the cluster solution. They do not occur in unique indecomposable solutions.

6 Numerical examples.

Example 6.1 (n-butane) N-butane (C_4H_{10}) was simulated at 300K for time-scale $\tau = 200fs$. For every configuration the torsion angles were stored which are sufficient for a rough reconstruction of the spatial coordinates together with the corresponding equilibrium bonds and angles. For conformational analysis like in [2] only the main torsion angle was considered and uniformly partitioned into 50 equidistant intervals $\Omega^* = \{A_1, \dots, A_{50}\}$. Since torsion angles between -30° and $+30^\circ$ did not occur during simulation, 8 intervals were skipped.

The corresponding Perron Cluster was $\lambda_1 = 1$, $\lambda_2 = 0.9906$ and $\lambda_3 = 0.9877$, while the 4th eigenvalue $\lambda_4 = 0.5234$ is bounded away from 1, i.e. $s = 3$.

Figure 1(a) shows an I_T -projection of Ω^* and its convex hull $co(\mathcal{C})$ which is indeed a 2-simplex. Hence the corresponding indecomposable solution is unique and the three membership functions are obtained via (LT2), see Figure 1(b).

The intervals $[57.6^\circ, 64.8^\circ]$, $[172.8^\circ, 180.0^\circ]$ and $[-64.8^\circ, -57.6^\circ]$ are the extreme elements of this clustering representing the trans- and the gauche-conformations of n-butane. The other intervals are more or less transition states between these conformations. There are no intermediate states.

Example 6.2 (VX-478) The inhibitor VX-478 of the enzyme HIV-protease was simulated at 700K. After simulation Ω was adaptively decomposed into 72 boxes $\Omega^* = \{A_1, \dots, A_{72}\}$ via self-organizing box maps (SOBM) [8]. T and its spectrum of eigenvalues were computed. Since the largest gap was between λ_2 and λ_3 , Perron Cluster analysis gave $s = 2$ and therefore $\theta = 0$. Note that there is a second large gap between λ_5 and λ_6 with relatively small $|\theta|$, but on low level with regard to λ_5 .

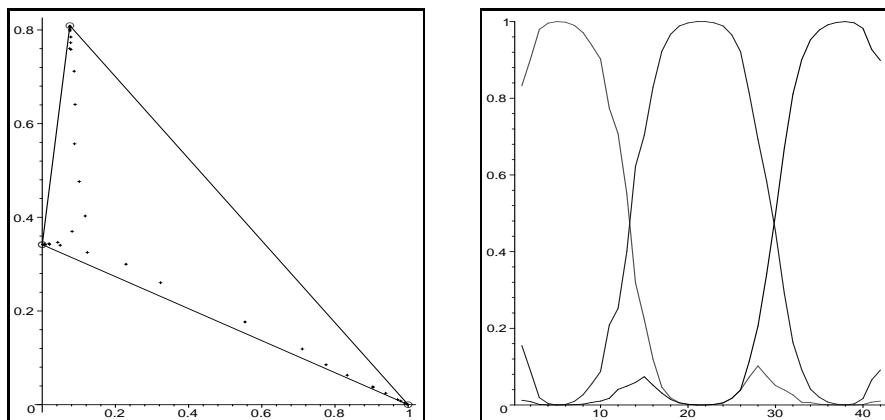


Figure 1: (a) I_T -projection of Ω^* for n-butane. (b) Indecomposable Solution $\{\chi_1, \chi_2, \chi_3\}$ for $m = 42$ and $s = 3$.

s	λ_s	$ \theta $
1	1.0000	
2	0.9302	0.0000
3	0.8850	0.8011
4	0.8758	0.4466
5	0.8602	0.1528
6	0.7948	0.6382

Table 1: Eigenvalues and indicator for VX-478 at 700K

s	λ_s	$ \theta $
1	1.0000	
2	0.9524	0.0000
3	0.9405	0.5605
4	0.9271	0.0856
5	0.9046	0.8215
6	0.8865	0.5990

Table 2: Eigenvalues and indicator for VX-478 at 500K

After separation of 2 clusters via Perron cluster analysis, VX-478 was simulated at 500K in one of these clusters. Again Ω was adaptively decomposed.

Table 2 gives us the eigenvalues and $|\theta|$ for the computed T . The gaps between λ_4 and λ_5 and between λ_5 and λ_6 are not significant enough to decide whether $s = 4$ or $s = 5$. But an investigation of $|\theta|$ shows that only $s = 4$ leads to an almost unique cluster solution.

7 Conclusion

This paper suggests a modification of conformational analysis. The crisp cluster problem is replaced by a fuzzy cluster problem. This leads to the following conceptual changes:

- Transition states are not assigned to one crisp cluster any longer. Instead grades of membership are computed via fuzzy membership functions yielding more information about the molecular dynamics.
- The introductory problem of assigning a transition state to one crisp cluster was responsible for the ambiguity of the cluster solution in classical set theory. Now by fuzzy set theory, an indecomposable solution of an almost T -invariant cluster problem can be unique although transition states occur.
- In many numerical examples there are no gaps in the spectrum of T , which are large enough to define the Perron Cluster. In this case the indicator for uniqueness can be used as a further hint to decide, how many clusters should be chosen.

Acknowledgments. It is a pleasure to thank Frank Cordes and Alexander Fischer for many helpful discussions.

References

- [1] P.Deuffhard W.Huisinga A.Fischer and Ch. Schütte. Identification of almost invariant aggregates in nearly uncoupled markov chains. *Linear Algebra and its Applications*, 315:39–59, 2000.
- [2] Ch.Schütte. *Conformational Dynamics: Modelling, Theory, Algorithm and Application to Biomolecules*. Habilitation Thesis, Dept. of mathematics and computer science, Free University Berlin, 1998.
- [3] C.Bradford Barber D.P.Dobkin and H.Huhdanpaa. The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software*, 22(4):469–483, 1996.
- [4] A.Fischer F.Cordes and Ch.Schütte. Hybrid monte carlo with adaptive temperature in mixed-canonical ensembles: Efficient conformational analysis of RNA. *J. Comp. Chem.*, 19(15):1689–1697, 1998.
- [5] M.Allen and D.Tildesley. *Computer Simulations of Liquids*. Clarendon Press, Oxford, 1990.
- [6] P.Deuffhard M.Dellnitz O.Junge and Ch.Schütte. *Computation of essential molecular dynamics by subdivision techniques*, in Computational Molecular Dynamics: Challenges, Methods, Ideas, edited by P.Deuffhard et al. Lecture Notes in Computational Science and Engineering, Springer-Verlag, Berlin, 1998.
- [7] R.Ostrovsky and Y.Rabani. *Polynomial time approximation schemes for geometric k -clustering*. Proceedings of the 41st Annual Symposium on Foundations of Computer Science, 2000.
- [8] T.Galliat P.Deuffhard R.Roitzsch and F.Cordes. *Automatic Identification of Metastable Conformations via Self-Organized Neural Networks*. Computational Methods for Macromolecules: Challenges and Applications, *Proceedings of 3rd International Workshop on Algorithms for Macromolecular Modelling*, New York, Oct. 12-14, 2000, T.Schlick and H.H.Gan, eds., *Lecture Notes in Computational Science and Engineering*, Springer-Verlag, 2002.
- [9] Ch.Schütte A.Fischer W.Huisinga and P.Deuffhard. A direct approach to conformational dynamics based on hybrid monte carlo. *Journal of Computational Physics*, 151:146–168, 1999.
- [10] H.-J. Zimmermann. *Fuzzy Set Theory - and Its Applications*. Kluwer Academic Publishers, 2nd ed., 1991.