Set-Oriented Dimension Reduction: Localizing Principal Component Analysis via Hidden Markov Models *

Illia Horenko[1], Johannes Schmidt-Ehrenberg[2] and Christof Schütte[1]

[1] Freie Universität Berlin, Department of Mathematics and Informatics,
Arnimallee 6, D-14195 Berlin, Germany
[2] Zuse Institute Berlin (ZIB)
Takustr. 7, D-14195 Berlin

Abstract. We present a method for simultaneous dimension reduction and metastability analysis of high dimensional time series. The approach is based on the combination of hidden Markov models (HMMs) and principal component analysis. We derive optimal estimators for the log-likelihood functional and employ the Expectation Maximization algorithm for its numerical optimization. We demonstrate the performance of the method on a generic 102-dimensional example, apply the new HMM-PCA algorithm to a molecular dynamics simulation of 12–alanine in water and interpret the results.

Introduction

Let us assume that the observation of the physical process under consideration (f. e. conformational dynamics of some biological molecule) is given in the form of a high dimensional time series in some molecular degrees of freedom (f. e. torsion angles or distances between some important groups of atoms in the molecule). The general task which arises in many practical applications is to find the few important or *essential* degrees of freedom that can explain most of the observed process and thus can help to understand the physical mechanism [1–4].

The increasing amount of "raw" simulation data and growing dimensionality of these simulations have led to a persistent demand for modeling approaches which allow to extract physically interpretable information out of the data. What is needed is *automatized* generation of low–dimensional physical models based on (noisy) data, i.e., interesting approaches should provide *data–based dimension reduction*. This should be carefully distinguished from analytical approaches like, e.g., the Zwanzig-Mori approach, the Karhunen-Loève expansion, or averaging techniques. The latter approaches allow to reduce the dimension of a given physical model, but the problem of finding essential coordinates must be solved previously and may be data–driven as well. See the textbook [5], or the excellent review article [6] for an overview. Compare also [7] for a related approach.

* Supported in part by the DFG Research Center MATHEON, Berlin.

The problem of dimension reduction becomes crucial when dealing with databases of molecular dynamics trajectories [8, 1]. Recent works show that even such simple linear dimension reduction strategies as principal component analysis (PCA) allow for a significant compression of the time–series information (factor 10 in [9]). However, such a linear technique as PCA applied to in general nonlinear phenomena as, e.g., transitions between the metastable conformations of biological molecules can be misleading and produce difficulties in the interpretation [1, 10, 8]. One way of trying to circumvent these problems is non–linear extension of PCA (NLPCA) [11]. However, this non-linear strategy is numerically expensive and not robust enough, thus resulting in restricted applicability of the technique [12]. Another possibility to extend the linear dimension reduction techniques is contained in the theory of the indexing of high dimensional databases, where the problem was partially solved by combining correlation analysis with clustering techniques [13–15]. But due to the fact that the proposed methods rely on *geometrical clustering* of possibly high dimensional data–spaces, the resulting algorithms rely on some sort of distance-metric and scale polynomially wrt. the length of the data-set. Alternatively, for the time series analysis of molecular dynamics trajectories, due to additional information encapsulated in the time component, it is possible to employ *dynamical clustering* techniques like hidden Markov models (HMMs) which scale linear wrt. the length of the time series [16–21].

In this paper we present a novel method for simultaneous dimension reduction *and* clustering of the time series into metastable states. The approach is based on the combination of the HMM with PCA. The problem of simultaneous dimension reduction and metastability analysis is solved by the optimization of an appropriate log–likelihood functional by means of the Expectation Maximization algorithm (EM) [18]. The performance of the resulting HMM–PCA algorithm is demonstrated by application to some model examples and to a microsecond simulation of 12–alanine protein in water.

1 Principal Component Analysis (PCA)

The simplest form of the dimension reduction is known in statistics as principle component analysis (PCA). Let the data be given in form of a sequence $\{x_t\}_{t=1,\dots,T}$ of states. The idea of the method consists in identification of m *principal* directions with highest variance in n-dimensional observed data $x_t : \mathbb{R}^1 \to \mathbb{R}^n$ ($m << n$). These directions are defined with the help of linear projectors $\mathbf{T} \in \mathbb{R}^{n\times m}$, i.e., $\mathbf{T}$ is understood to project onto the subspace spanned by the principal directions. Mathematically the problem of identifying $\mathbf{T}$ can be stated as minimization of the residuum–functional (which describes the least–squares difference between the original observation and its reconstruction by means of the m-dimensional projection):

$$\mathbf{L}(x_t, \mathbf{T}, \mu) = \sum_{t=1}^{T} \left\| (x_t - \mu) - \mathbf{T}\mathbf{T}^{\mathsf{T}}(t)\,(x_t - \mu) \right\|_2^2. \tag{1}$$

The functional $\mathbf{L}$ depends on the projector matrices $\mathbf{T}$ and *center vectors* $\mu \in \mathbb{R}^n$. Moreover, the projectors $\mathbf{T}$ are subjected to the orthogonality condition:

$$\mathbf{T}^{\mathsf{T}}\mathbf{T} = Id^{m\times m}, \tag{2}$$

The functional (1) can be equivalently written as

$$\mathbf{L} = \sum_{t=1}^{T} (x_t - \mu)^{\mathsf{T}} \left(Id - \mathbf{T}\mathbf{T}^{\mathsf{T}} \right) (x_t - \mu) \,. \tag{3}$$

This functional can be minimized analytically subject to the orthogonality conditions resulting in the expressions for optimal parameters μ and $\mathbf{T}$:

$$\mu = \frac{1}{T}\sum_{t=1}^{T} x_t, \quad C = \frac{1}{T}\sum_{t=1}^{T}(x_t - \mu)(x_t - \mu)^{\mathsf{T}} = \mathbf{T}S\mathbf{T}^{\mathsf{T}} + \mathcal{O}(\Lambda_{min}) \tag{4}$$

where $\mathbf{T}$ is the matrix of m dominant eigenvectors and $S = diag(\lambda_1, .., \lambda_m)$ contains m corresponding largest eigenvalues of the covariance matrix C. $\Lambda_{min} = diag(\lambda_{m+1}, \ldots, \lambda_n)$ is the diagonal matrix containing the remaining part of the spectrum. This result means that the optimal value of the parameter μ is given simply by the expectation value of the data and the corresponding optimal projector $\mathbf{T}$ is defined by the dominant eigenvectors of the data covariance–matrix. It is important to mention that nowhere in the derivation of the optimal estimator the assumption about the form of the x_t distribution is needed.

However, in many interesting cases the standard PCA–approach does not result in a meaningful dimension reduction. Let us assume, for example, that the time series given results from a realization of the process governed by a two–dimensional double-well Langevin–dynamics of the form

$$\ddot{x}(t) = -\mathrm{grad}\, V\left(x(t)\right) - \gamma\dot{x}(t) - \sigma\dot{W}(t), \tag{5}$$

with friction matrix $\gamma = \begin{pmatrix} 0.25 & 0.125 \\ 0.125 & 0.25 \end{pmatrix}$, noise matrix $\sigma = \begin{pmatrix} 0.6 & 0 \\ 0 & 0.6 \end{pmatrix}$ and potential energy defined as the sum of two Gaussian wells orthogonal to each other added to a harmonic potential:

$$\begin{aligned} V(x) = &\sum_{l=1}^{2} a_l \exp\left(-(x - \mu_l^{sys})^{\mathsf{T}} D_l^{sys} (x - \mu_l^{sys}) \right) \\ &+6(x - 0.5(\mu_l^{sys} + \mu_l^{sys}))^{\mathsf{T}} (D_1^{sys} + D_2^{sys})(x - 0.5(\mu_l^{sys} + \mu_l^{sys})), \end{aligned} \tag{6}$$

$$D_1^{sys} = \begin{pmatrix} 20 & 0 \\ 0 & 0.5 \end{pmatrix}, \quad D_2^{sys} = \begin{pmatrix} 0.5 & 0 \\ 0 & 20 \end{pmatrix}, \quad \mu_1^{sys} = \begin{pmatrix} 0 \\ 1.5 \end{pmatrix}, \quad \mu_2^{sys} = \begin{pmatrix} 1.5 \\ 0 \end{pmatrix}$$

The Langevin dynamics in this case produces two clusters of states each associated with the corresponding *metastable* well. The application of PCA via (4) with $m = 1$ to this time series results in an inadequately reconstructed dynamics. If we first cluster the time series into two clusters and then apply

(4) to each of the clusters separately, we can reduce the value of the residuum–functional (1) from 191.1 in a "global" PCA case to 46.3 in a "local" one (by "local" PCA we understand the PCA for each of the clusters, the value of the residuum-functional is then given by the sum of the "local" functionals). This also results in a much better quality of the data-reconstruction.

This leads us to a simple idea: if we want to enhance the performance of PCA–based dimension reduction we should exploit the internal structure of the data, i.e., we should decompose the time series of the observed process into metastable aggregates and then make the "local" dimension reduction by means of PCA. Furthermore, we can state more ambitious question: Is it probably possible to use the local principle dimensions as tokens in the clustering of the time series itself. If it is possible this will allow to combine clustering of data and dimension reduction in one algorithmic step hopefully leading to synergetic effects and allowing both clustering of the time series in metastable sets and the dimension reduction.

1.1 Hidden Markov Models (HMM)

A hidden Markov model (HMM) is a stochastic process with hidden and observable states. The hidden process consists of a sequence $X_1, X_2, X_3, \ldots$ of random variables taking values in some "state space", the value of X_t being "the state of the system at time t". In applications these states are not observable, and therefore called *hidden*. Each state causes a specific output that might be either discrete or continuous. This output is distributed according to a certain conditional distribution (conditioned to the hidden state). Thus, realizations of HMM are concerned with two sequences, an observation sequence and a sequence of hidden states.

The dynamics under consideration is assumed to be a Markov process, that is, the state sequence has the Markov property which means that the conditional distribution of the "future" X_{n+1} given the "past", $X_1, \ldots, X_n$, depends on the past only through X_n. Since the HMM state space in general is finite, we thus are concerned with a Markov chain, which is characterized by the so-called transition matrix, whose entries correspond to the probabilities of switching from one state to another. The sum of all coefficients in one row is the probability of taking any state, therefore being one, which means that the transition matrix is a row-stochastic matrix.

A HMM is designed to describe the situation in which part of the information of the system is unknown (or hidden) and another part is observed. In molecular dynamics the information initially is hidden in which metastable subset (conformation) the molecular system is at a certain instant in time, while the information on the state of the selected torsion or backbone angles is completely known. A HMM then consists of a Markov chain model for the hidden (metastable) states that encodes with which probability one switches from one hidden state to another, and a *conditional probability* of observation of specific torsion angles *if* one is in a certain hidden state. To describe the whole system, we need to know the number of hidden states, the transition matrix between

them, an initial distribution, and for each state a certain probability distribution for the observation.

Therefore, a HMM formally is defined as a tuple $\lambda = (S, V, A, B, \pi)$ where

- $S = \{s_1, s_2, ..., s_L\}$ is a set of a finite number L of states,
- $V \subset \mathbb{R}^k$ is the observation space,
- $A = (a_{ij})$ is the transition matrix, where $a_{ij} = P(X_{t+1} = s_j | X_t = s_i)$ describes the the transition probability from state s_i to state s_j,
- B_k, $k = 1, \ldots, N$ are probability density functions in the observation space,
- $\pi = \pi_i$ is a stochastic vector, that describes the initial state distribution, $\pi_i = P(X_1 = s_i)$.

Often, the short notation $\lambda = (A, B, \pi)$ is used since S and V are implicitly included. HMMs can be set up for discrete or continuous observations. For continuous observations the most popular choice is to use (multivariate) normal distributions for the output distributions B_k.

1.2 HMM-PCA

The fitting of the parameters can be pefomred with the help of the *maximum likelihood principle*. The likelihood function is $L(\lambda) = P(x_t, X_t|\lambda)$, i.e., we consider the observation sequence as being given and ask for the variation of the probability in terms of the parameters. The maximum likelihood principle then simply states, that the optimal parameters are given by the absolute maximum of L. Thus, similarly to the PCA dimension reduction, the maximum likelihood principle is an optimization problem in parameter space.

In order to combine both approaches, we first make two assumptions on the observation process: (i) the observed data in the hidden states are distributed according to a multivariate Gaussian distributions ρ_B, (ii) the hidden process switching between the metastable states is Markovian, i.e., the probability of the conformational change depends on the current conformation only. The first assumption is approximately valid for a large class of observables in molecular dynamics (f. e. for the torsion angles or chemical bond lengths in the molecule). The second assumption is connected to the characteristic timescale at which the memory kernel of the molecular system is decaying and is also satisfied for a wide class of applications.

These assumptions allow to design a statistical model for the observed data and to construct the likelihood function for a reduced system. In analogy to the residuum–functional (3) we have

$$P(x_t, X_t|\lambda) = \pi_{X_0} e^{-\frac{1}{2}\left(x_0-\mu_{X_0}\right)^\top \mathbf{T}_{X_0} S_{X_0} \mathbf{T}_{X_0}^\top \left(x_0-\mu_{X_0}\right)} \prod_{k=1}^{T-1} \frac{A(X_k, X_{k+1})}{\sqrt{(2\pi)^m \det(S_{X_{k+1}})}} \times e^{-\frac{1}{2}\left(x_{k+1}-\mu_{X_{k+1}}\right)^\top \mathbf{T}_{X_{k+1}} S_{X_{k+1}} \mathbf{T}_{X_{k+1}}^\top \left(x_{k+1}-\mu_{X_{k+1}}\right)} \tag{7}$$

where $B_i = (\mu_i, \mathbf{T}_i, S_i)$ is a set of multivariate Gaussian distribution parameters where $\mu_i \in \mathbb{R}^n$ are the centers of the clusters, $\mathbf{T}_i \in \mathbb{R}^{n \times m}$ the corresponding

optimal projectors, and $S_i \in \mathbb{R}^{m\times m}$ a diagonal matrix of dominant variances. Functional (7) should be additionally subjected to constraints: (i) the projector orthogonality condition (2),(ii) and the condition for stochasticity of the transition matrix A (i.e., the row sums of the matrix should be 1.0).

For numerical reasons it is much more convenient to take the logarithm of the likelihood functional and optimize the resulting log-likelihood functional. Writing the log-likelihood together with both constraints in Lagrange–form, taking the derivatives wrt. to the model parameters and setting them to zero we get:

$$\mu_i = \frac{1}{T}\sum_{t=1}^{T}\alpha_i(t)\beta_i(t)x_t, \quad \sum_{t=1}^{T}\alpha_i(t)\beta_i(t)(x_t-\mu)(x_t-\mu)^{\mathsf{T}} = \mathbf{T}_i S_i \mathbf{T}_i^{\mathsf{T}} + \mathcal{O}(\Lambda_{min}^i)$$

where $\alpha_i(t), \beta_i(t)$ are forward and backward variables (as usually defined in the context of HMMs, see [21]). They are related to the Markov process (A, π) and describe the probabilities to observe the hidden process X_t in the state i in the time t. We observe direct correspondence between the estimator formulas (8) and those given by standard PCA (4). That is, in the case of a single hidden state the minimization of the HMM-PCA functional (7) is equivalent to optimization of the residuum–functional (3). Only the dominant eigenvectors are needed for the construction of matrix $\mathbf{T}$. One can compute them efficiently with some iterative subspace method (e.g. Lanczos, cf. [22]). In the case of several hidden states we suggest to use the standard Expectation-Maximization algorithm [23], often also called the Baum-Welch algorithm [16, 17]. The Expectation-Maximization (EM) algorithm is a maximum likelihood approach that improves iteratively an initial parameter set, and converges to a local maximum of the likelihood function. Its two steps, the E- and M-steps, are iteratively repeated until the improvement of the likelihood becomes smaller than a given limit. In all other details the EM algorithm used herein follows standard procedures.

To apply the EM algorithm to a given observation sequence, we have to set up a HMM $\lambda = \lambda(A, B, \pi)$ by assuming a finite number L of hidden states, a distribution function for the output of each state, and an initial values for all remaining parameters.

2 Numerical Examples

2.1 Langevin Dynamics in 102 Dimensions

As a first example we consider realizations of the Langevin equation (5) with $x = (q, y) \in \mathbb{R}^2 \times \mathbb{R}^N$ and the perturbed two–hole potential

$$V(x) = \sum_{l=1}^{2} a_l \exp\left(-(q-\mu_l^{sys})^{\mathsf{T}} D_l^{sys}(q-\mu_l^{sys})\right) + \frac{1}{2}y^{\mathsf{T}} D_{bath} y \tag{8}$$

$$+\delta_0\Big(\cos(2\pi k(x_1+x_2)) + \cos(2\pi k(x_1-x_2))\Big), \tag{9}$$

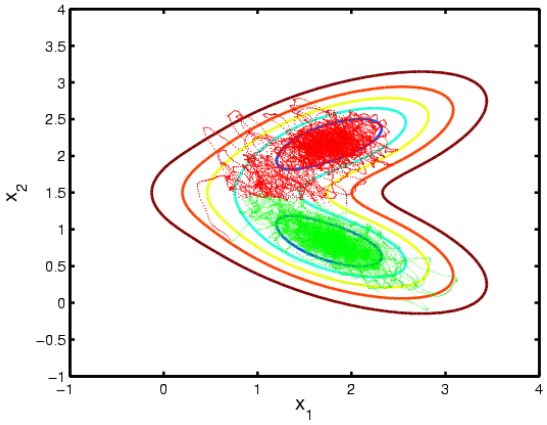

Fig. 1. Clustering of the time series resulting from the application of the Viterbi-algorithm [24]. The 102–dimensional data is back–rotated and projected on metastable dimensions.

where $\delta_0 \ll 1$ is a small perturbation parameter. The N harmonic bath variables are denoted by y, whereas x labels the two "metastable" dimensions that live in the plane of the double well potential. We have chosen the following parameter values: $\mu_1^{sys} = (1.8, 2.2)^\mathsf{T}$, $\mu_2^{sys} = (1.8, 0.8)^\mathsf{T}$, $D_1^{sys} = \begin{pmatrix} 1 & -1 \\ -1 & 3 \end{pmatrix}$, $D_2^{sys} = \begin{pmatrix} 1 & 1 \\ 1 & 3 \end{pmatrix}$, $a_1 = -6$, $a_2 = -6$ such that we get two contiguously placed skew wells and make identification of the metastable sets more challenging compared to a well–separated situation. The parameter matrices D_{bath} and γ have been chosen to be symmetric, positive definite, and tri-diagonal, with 10.0 on the main diagonal and 5.0 on secondary diagonals for D_{bath} (5.0 and 2.5 respectively for γ). The noise parameter σ was taken as a diagonal matrix with 4.0 on the diagonal. The system is metastable because the barrier is sufficiently larger than the average kinetic energy in the system.

Simulation of the model has been realized with the Euler–Maruyama integrator (discretization time step $\Delta t_{Euler} = 0.0002$) and total time length 500. Each hundredth instance of the resulting time series has been taken for a subsequent parameter estimation (resulting in observation time step $\tau = 0.02$) such that $T = 25.000$.

Furthermore, in order to make our model system more realistic and mimic the features inherent in biological systems, we rotate the resulting time series in the (N+2) dimensional space. We do it in such a way, that the metastability of the system becomes *hidden* in all the dimensions of the system. Application of the HMM–PCA method indicates the presence of two metastable states in the time series. In order to interpret the quality of the resulting model, we rotate the time series back, color the elements according to the corresponding metastable state and plot them atop of the original potential surface in (x_1, x_2). As we can

see in Fig. 1, the local Langevin models are correctly situated at the wells of the double–well potential in the metastable dimensions and the elements of the time series are assigned in a proper way.

2.2 Analysis of the Long–Time Behavior of 12-Alanine in Water.

As a second application we analyzed a molecular dynamics trajectory of 12-Alanine for conformational changes. The molecule was simulated using CHARMM with an implicit water environment. The data were kindly provided by Jeremy Smith and Frank Noé (IWR Heidelberg). We analyzed a $1\mu s$ long simulation with $2fs$ time steps. The analysis was performed on the basis of the 33 backbone torsion angles. We identified 3 metastable states that were analyzed with $m = 3$ local PCA modes. The left column of Fig. 2 illustrates the shape variability of the three conformations using a technique from [25]. The geometries of all time steps are accumulated into a pseudo density that is visualized using direct volume rendering. In the right column, the mean backbone shapes of the three states are depicted together with arrows that indicate the dominant PCA-dimensions inside the conformations.

The identified metastable sets are the conformations of the molecule [26], which have life times of 10 ns. As we can see, the dominant dimensions can be interpreted as principal movements of the backbone which are characteristic for the corresponding conformations. The analysis shows that in terms of HMM-PCA these characteristic movements can be used to distinguish between the metastable sets and can be helpful in the clustering of the time series.

3 Conclusion

We presented a novel HMM based method for simultaneous dimension reduction and clustering of time series data. The method is based on a combination of an HMM approach and local PCA analysis. Incorporation of the local PCA analysis helps to map the clustering problem into low dimensional space. We have demonstrated the application of the method for a model system and a molecular dynamics trajectory of 12-Alanine. The numerical examples demonstrate the usefulness of the HMM-PCA approach.

References

1. T. Ichiye and M. Karplus. Collective motions in proteins – a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins*, 11:205–217, 1991.
2. D. Frenkel and B Smit. *Understanding Molecular Dynamics: From Algorithms to Applications*. Academic Press, London, 2002.
3. W. E and E. Vanden-Eijnden. Metastability, conformation dynamics, and transition pathways in complex systems. In S. Attinger and P. Koumoutsakos, editors, *Multiscale, Modelling, and Simulation*, pages 35–68. Springer, Berlin, 2004.

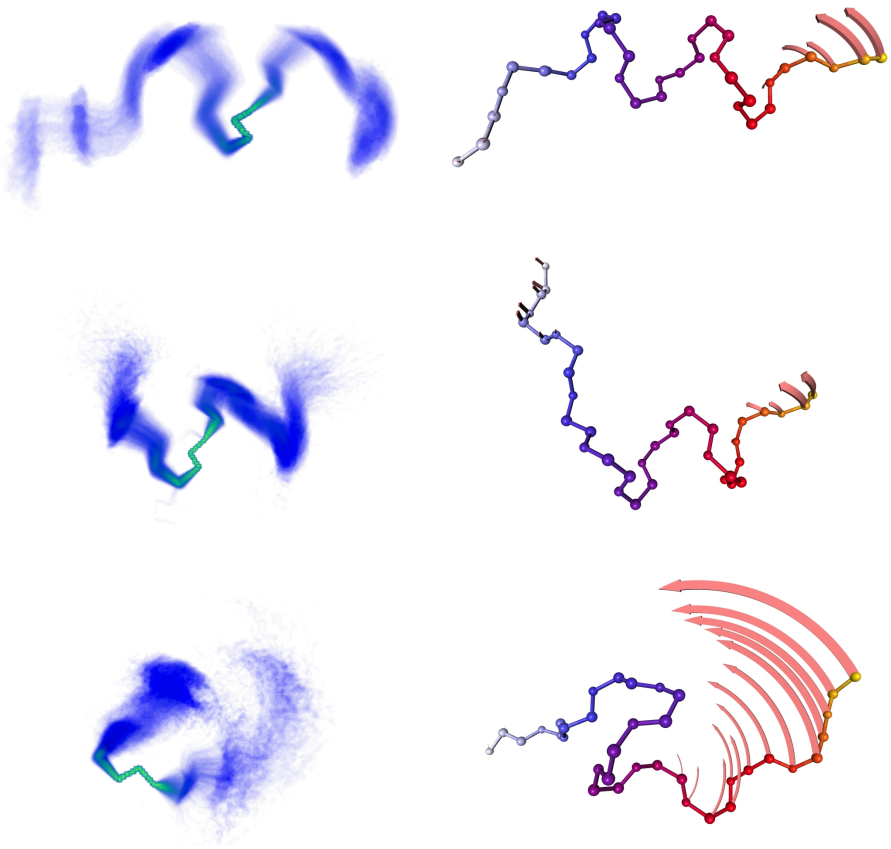

Fig. 2. *Three dynamic conformations of 12-Alanine. The left column shows direct volume renderings of pseudo densities corresponding to the conformations in the investigated trajectory. The right column indicates the flexibilities of the corresponding HMM-PCA-modes with highest variance. Images on the left show flexibility at both ends of the backbone, while images on the right indicate a concentration on one end. The flexibility of the other end is covered by the respective second PCA-modes, which are not shown here.*

4. P. Deuflhard and C. Schütte. Molecular conformation dynamics and computational drug design. In *Applied Mathematics Entering the 21st Century: Invited Talks from the ICIAM 2003 Congress*, 2004.
5. P. Holmes, J.L. Lumley, and G. Berkooz. *Turbulence, Coherent Structures, Dynamical Systems and Symmetry.* Cambridge University Press, 1996.
6. D. Givon, R. Kupferman, and A. Stuart. Extracting macroscopic dynamics: Model problems and algorithms. *Nonlinearity*, 17:R55–R127, 2004.

7. R. Kupferman and A.M. Stuart. Fitting sde models to nonlinear kac-zwanzig heat bath models. *Physica D*, 199:279–316, 2004.
8. M.A. Balsera, W. Wriggers, Y. Oono, and K. Schulten. Pricipal Component Analysis and long time protein dynamics. *J. Chem. Phys.*, 100:2567–2572, 1996.
9. T. Meyer, C. Ferrer-Costa, A. Perez, M. Rueda, A. Bidon-Chanal, F. Luque, C. Laughton, and M. Orozco. Essential dynamics: a tool for efficient trajectory compression and management. *JCTC*, 2:251–258, 2006.
10. P.H. Hünenberger, A.E. Mark, and W.F. van Gunsteren. Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations. *J. Mol. Biol.*, 252:492–503, 1995.
11. A.H. Monahan. Nonlinear principal component analysis by neural networks: Theory and application to the lorenz system. *J. Climate*, 13:821–835, 2000.
12. B. Christiansen. The shortcomings of NLPCA in identifying circulation regimes. *J. Climate*, 18:4814–4823, 2005.
13. C. Agarrval, J. Wolf, P. Yu, C. Procopiuc, and J. Park. Fast algorithms for projected clustering. *Proceedings of the 1999 ACM SIGMOD international conference on Management of data*, 1999.
14. K. Chakrabarti and S. Mehrotra. Local dimensionality reduction: A new approach to indexing high dimensional spaces. In *Proceedings of the 26th VLDB Conference*, pages 98–115. Cairo,Egypt, 2000.
15. P. Zhang, Y. Huang, S. Shekhar, and V. Kumar. Correlation analysis of spatial time series datasets: A filter-and-refine approach. *the Proc. of the 7th Seventh Pacific-Asia Conference on Knowledge Discovery and Data Mining(PAKDD 2003)*, 2003.
16. L.E. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occuring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Stat.*, 41:164–171, 1970.
17. L.E. Baum. An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. *Inequalities*, 3:1–8, 1972.
18. J.A. Bilmes. *A Gentle Tutorial of the EM Algorithm and its Applications to Parameter Estimation for Gaussian Mixture and Hidden Markov Models. Thechnical Report.* International Computer Science Institute, Berkeley, 1998.
19. Z. Ghahramani. An introduction to hidden Markov models and Bayesian networks. *Int. J. Pattern Recognition and Artificial Intelligence*, 15(1):9–42, 2001.
20. J. Frydman and P. Lakner. Maximum likelihood estimation of hidden Markov processes. *Ann. Appl. Prob.*, 13(4):1296–1312, 2003.
21. I. Horenko, E. Dittmer, A. Fischer, and C. Schütte. Automated model reduction for complex systems exhibiting metastability. *SIAM Multiscale Modeling and Simulation*, 2005. accepted for publication.
22. G.H. Golub and C.F. van Loan. *Matrix computations*. The John Hopkins University Press, second edition edition, 1989.
23. A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc. B*, 39:1–38, 1977.
24. A.J. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Trans. Informat. Theory*, 13:260–269, 1967.
25. Johannes Schmidt-Ehrenberg, Daniel Baum, and Hans-Christian Hege. Visualizing dynamic molecular conformations. In *Proceedings of IEEE Visualization 2002*, pages 235–242, 2002.
26. Christof Schütte, Alexander Fischer, Wilhelm Huisinga, and Peter Deuflhard. A direct approach to conformational dynamics based on hybrid Monte Carlo. *J. Comput. Phys.*, 151:146–168, 1999.