Noname manuscript No.
(will be inserted by the editor)

Dimension Reduction for Time Series with Hidden Phase Transitions and Economic Applications*

Illia Horenko[1], **Christof Schütte**[2]

[1] Institut für Mathematik II, Freie Universität Berlin, Arnimallee 2-6, 14195 Berlin, Germany, e-mail: horenko@math.fu-berlin.de
[2] Institut für Mathematik II, Freie Universität Berlin, Arnimallee 2-6, 14195 Berlin, Germany, e-mail: schuette@math.fu-berlin.de

The date of receipt and acceptance will be inserted by the editor

Abstract We present a descriptive method for simultaneous dimension reduction and identification of temporal phases in high dimensional time series. The aim is to find hidden states in the multidimensional time series data and to identify the essential dimensions for each of the hidden states without making assumptions about the type of the underlying essential dynamics and structure of the noise process (additive or multiplicative). We solve this problem by optimization of an appropriate functional and demonstrate two different numerical approaches: (i) a direct method which does not imply assumptions on neither the hidden or observed processes, and (ii) a method based on the combination of hidden Markov models (HMMs) and principal component analysis (PCA) which assumes Markovian dynamics of the hidden process and multivariate Gaussian distributions for observed variables. We derive optimal estimators for the log-likelihood functional and employ the Expectation– Maximization algorithm for its numerical optimization. We demonstrate the performance of presented approach on a generic multi–dimensional model system and some financial time series. We show how the method can help to find hidden phases, to extract the leading market indicators and to analyze the principal correlation patterns hidden in the data.

Introduction

Assume that an observation of a process under consideration is given in the form of a high dimensional time series in some degrees of freedom. The

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

general task 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 underlying mechanism [1–5].

The increasing amount of "raw" measurement data and growing dimensionality of investigated processes have led to a persistent demand for modelling approaches which allow to extract interpretable information out of the data. What is needed, is *automatized* generation of low–dimensional models based on (noisy) data, i.e., there is a demand for approaches which provide *data–based dimension reduction*. This should be carefully distinguished from analytical approaches. The latter approaches allow to reduce the dimension of a given model, but the problem of finding essential coordinates must be solved previously and may be data–driven as well. See the textbook [6], or the excellent review article [7] for an overview.

The problem of dimension reduction becomes crucial when dealing with high-dimensional time-series or large data-sets. It has been shown that even such simple linear dimension reduction strategies as principal component analysis (PCA, also known as Karhunen-Loève expansion or proper orthogonal decomposition) allow for a significant compression of the time–series information, cf. [8]. However, such a linear technique as PCA applied to in general nonlinear phenomena as, e.g., transitions between different phases or regimes of a multidimensional dynamical system can be misleading and can produce difficulties in the interpretation [1,9,10]. One way to circumvent these problems is non–linear extension of PCA (NLPCA) [11]. However, this non-linear strategy is numerically expensive and not robust enough, so the applicability of the technique is restricted [12]. Another possibility to extend the linear dimension reduction techniques is contained in the theory of indexing of high dimensional data-bases, where the problem was tackled 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 therefore scale polynomially wrt. the length of the data-set. An alternative, as proposed in this paper, is to employ *dynamical clustering* techniques like hidden Markov models (HMMs) which scale linear wrt. the length of the time series [16–22] and to combine them with the localized version of PCA.

A point of crucial importance in construction of reduced models based on observation data is the number and type of dynamical structural assumptions being used in the modelling of the underlying system. If the "physics" of the system is known (i. e. if the mathematically rigorous model can be analytically deduced from the "first principles"), then it is possible to construct what one can call an *explanatory* model of the dynamics. This *dynamical model* can then be further parameterized using an appropriately constructed estimator. If this is not the case and the proper dynamical model from "first principles" is not known a priory, there are two possible ways to describe the data: (i) one can try to fit a dynamical model from a known class of models (f. e. autoregressive integrable moving average

(ARIMA, [23]), multivariate autoregressive models (MVAR, [23]), conditional heteroscedastic models (ARCH, [24]) or their generalizations) to the observed data-set , (ii) one can use some kind of *information criterion* to find out the essential dimensions (f. e. in the sense of the lowest reconstruction error when the full data-set is reconstructed from its part). Approach (i) will allow to make a data-based parameterization of the chosen model but there is no guarantee that (a) the same parameterization would be valid for other observations of the same system and (b) that there exist no better dynamic model to describe the data. The goal of the approach (ii) is to find spatial patterns in the data, whereas the goal of (i) is to find dynamical patterns of certain predefined type.

Many economic applications are characterized by the need to find low-dimensional models for complex systems that undergo transitions between different temporal phases or "regimes". The information about these phases and the paths describing the movement of the system through them are hidden in the multi–dimensional observation data [17,25,26]. Dimension reduction approaches known in finance and econometrics literature are represented by methods of canonical correlation analysis [27–30], dynamical factor models based on principal component analysis [31–34], hidden Markov approaches for multivariate autoregressive models (VAR) [35–37] and other related approaches. Due to the fact that the application of the econometrics models almost always implies the construction of predictions based on historical observation data, all of the above approaches share the same feature - they are *explanatory*, i. e., the multidimensional dynamical model (usually it is a multivariate autoregressive model or one of its modifications) is mandatory for dimension reduction.

There are some recent indications that the whole variety of existing explanatory models fails to reproduce the intraday returns on the stock markets [38]. This particular point motivates the presented work. In this paper we present a novel *descriptive method* for simultaneous dimension reduction *and* clustering of the time series into metastable states. We derive the expression for the residuum–functional; minimization of this functional allows to cluster the time series according to the differences in essential dimensions without making additional assumptions about type of the dynamical model in each of the hidden states. We present two alternative numerical methods to perform the optimization of the presented functional under following two conditions: (1) We have to assume that there are not too many regimes and that these are "metastable", i.e., the system typically remains within a certain regime for a time span that is significantly longer than the lag time of the time series. (2) Within each regime one is interested in the subspace that contains most of the variability of the time series. We will present two different numerical schemes. The first approach is based on the direct optimization of the functional. This approach involves the solution of some high-dimensional optimization problem and is therefore numerically expensive. The second approach is based on the combination of HMMs with PCA. The problem of simultaneous dimension reduction and metastability

analysis in that case is solved by the optimization of an appropriate log–likelihood functional by means of the Expectation Maximization algorithm (EM) [19]. This method is numerically favorable, since it performs linearly in the length of the observation sequence and so far can be used for very long observation sequences. However, the applicability of this method is restricted to the cases when the observation process in each of the states is multivariate Gaussian and the hidden process is Markovian. The performance of both methods is demonstrated by application to 202–dimensional model example and to the time series originating from daily observations of important stock market titles.

1 State-specific Principal Component Analysis

Let the data be given in form of a sequence $\{z_t\}_{t=1,\dots,T}$ of c-dimensional data vectors which describe the observation or measurement of a process at T consequent instances. In order to encounter for memory effects in the analyzed data (i. e. to be able to describe causal temporal dependencies existent in the observed sequence), we can extend the vector space of observables z_t at each time t with k previous observations $\{z_{t-1},\dots,z_{t-d}\}$ [39, 40]. The resulting vector $x_t = \{z_t, z_{t-1},\dots,z_{t-d}\}$ is a component in $n = dc$-dimensional space. The idea of the method is to identify the m *principal directions* with the highest variance in n-dimensional data x_t ($m << n$). In contrast to standard PCA, where this principal directions are supposed to be *global* (i. e. valid for the whole time series x_t), we will construct a *local or phase-specific* variant of PCA. We assume that principal directions can vary in time and are defined with the help of a sequence of K linear projectors $\mathbf{T}_i \in \mathbb{R}^{n\times m}, i = 1,\dots,K$, i.e., $\mathbf{T}_i$ is understood to project onto the subspace spanned by the local principal directions. Mathematically the problem of identifying $\mathbf{T}_i$ can be stated as a minimization problem wrt. the residuum–functional, describing the least–squares difference between the original observation and its reconstruction by means of the m-dimensional projection:

$$\mathbf{L}(x_t, \mathbf{T}_i, \mu_i) = \sum_{i=1}^{K}\sum_{t=1}^{T} \gamma_i(t) \left\| (x_t - \mu_i) - \mathbf{T}_i \mathbf{T}_i^{\mathsf{T}} (x_t - \mu_i) \right\|_2^2, \tag{1}$$

where $\gamma_i(t)$ denotes the probability to optimally describe the n-dimensional vector x_t at time t with the local projector $\mathbf{T}_i$ and $\sum_{i=1}^{K} \gamma_i(t) = 1$ for all t. This number gives a relative weight to an observation x_t in the hidden state i, similarly to what was proposed by [41] in the context of signal extraction and filtering. For the moment we assume the sequence of probabilities $\gamma_i(t)$ to be known and fixed, in the next section we will present a way to estimate this sequence from a given observation x_t. The functional $\mathbf{L}$ depends on the projector matrices $\mathbf{T}_i$ and *center vectors* $\mu_i \in \mathbb{R}^n$. Moreover, the projectors

$\mathbf{T}_i$ are subjected to the orthogonality condition:

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

The functional (1) can be equivalently written as

$$\mathbf{L} = \sum_{i=1}^{K}\sum_{t=1}^{T}\gamma_i(t)(x_t-\mu_i)^{\mathsf{T}}Q_i\left(x_t-\mu_i\right), \tag{3}$$

where $Q_i = Id^{n\times n} - \mathbf{T}_i\mathbf{T}_i^{\mathsf{T}}$ is the projector onto the orthogonal complement of the range of $\mathbf{T}_i$.

Applying the Lagrange form of the functional with constraints we get

$$\mathbf{L} = \sum_{i=1}^{K}\sum_{t=1}^{T}\gamma_i(t)(x_t-\mu_i)^{\mathsf{T}}Q_i(x_t-\mu_q) + \sum_{i=1}^{K} e^T\Lambda_i \bullet \left(Id^{m\times m} - \mathbf{T}_i^{\mathsf{T}}\mathbf{T}_i\right)e, \tag{4}$$

where Λ_i is a symmetric matrix of the Lagrange multipliers (since (2) results in a symmetric matrix of constraints), $e = (1,1,...,1) \in \mathbb{R}^m$, and $A \bullet B$ denotes entry-wise multiplication of the matrices A and B. While the derivative wrt. Λ_i gives back the orthogonality condition (2), the derivatives of $\mathbf{L}$ with respect to μ_i and $\mathbf{T}_i$ yield the two conditions

$$\begin{aligned}
\frac{\partial\mathbf{L}}{\partial\mu_i} &= 2\sum_{t=1}^{T}\gamma_i(t)\left((x_t-\mu_i) - \mathbf{T}_i\mathbf{T}_i^{\mathsf{T}}(x_t-\mu_i)\right) = 0 \\
\frac{\partial\mathbf{L}}{\partial\mathbf{T}_i} &= -2\sum_{t=1}^{T}\gamma_i(t)\Big((x_t-\mu_i)(x_t-\mu_i)^T\mathbf{T}_i\Big) - 2\mathbf{T}_i\Lambda_i = 0,
\end{aligned} \tag{5}$$

for minima of $\mathbf{L}$. These conditions yield the following first results on the minimizers μ_i and $\mathbf{T}_i$:

$$Q_i(\mu_i-\bar{x}_i) = 0, \text{ with } \bar{x}_i = \frac{\sum_{t=1}^{T}\gamma_i(t)x_t}{\sum_{t=1}^{T}\gamma_i(t)} \quad\Rightarrow\quad \mu_i = \bar{x}_i + \xi_i, \tag{6}$$

where ξ_i is an arbitrary vector from the kernel of Q_i, i.e., $Q_i\xi_i = 0$, and

$$C_i\mathbf{T}_i = \sum_{t=1}^{T}\gamma_i(t)(x_t-\mu_i)(x_t-\mu_i)^{\mathsf{T}}\mathbf{T}_i = -\mathbf{T}_i\Lambda_i \tag{7}$$

with C_i being the weighted covariance matrix of the data in the hidden state i.

Obviously for $m = 1$ the last equation implies that $\mathbf{T}$ is an eigenvector of the matrix C_i (corresponding to an eigenvalue λ_i of C_i such that $\Lambda_i = -\lambda_i$). Via induction in m and by using the orthogonality constraint (2) together with the symmetry of Λ_i it is easy to show that the solution of (7) satisfies $\Lambda_i = -\lambda_i$ where λ_i is a diagonal matrix of eigenvalues of the covariance matrix C_i, and the columns of $\mathbf{T}_i$ are given by the corresponding (normalized) eigenvectors, cf. [42] for the global variant of PCA.

In the following we assume that C_i has at least m non–negative eigenvalues. In order to identify the proper m eigenvalues and eigenvectors that minimize $\mathbf{L}$, we first note that

$$(x_t-\mu_i)^{\mathsf{T}}\mathbf{T}_i\mathbf{T}_i^{\mathsf{T}}(x_t-\mu_i)=\mathrm{tr}\Big[\mathbf{T}_i^{\mathsf{T}}(x_t-\mu_i)(x_t-\mu_i)^{\mathsf{T}}\mathbf{T}_i\Big]. \tag{8}$$

Inserting (8) into the residuum-functional (4) and making use of the linearity of the trace operation we get

$$\mathbf{L}=\sum_{i=1}^{K}\Big[\sum_{t=1}^{T}\gamma_i(t)(x_t-\mu_i)^2-\gamma_i(t)\,\mathrm{tr}\Big[\mathbf{T}_i^{\mathsf{T}}C_i\mathbf{T}_i\Big]+e^T\lambda_i\bullet\big(Id^{m\times m}-\mathbf{T}_i^{\mathsf{T}}\mathbf{T}_i\big)\,e\Big]. \tag{9}$$

This result means that the minimization of the *least-squares residuum-functional* (4) with respect to $\mathbf{T}_i$ is equivalent to constraint maximization of the *variation functional*

$$\mathcal{L}=\sum_{i=1}^{K}\Big[\mathrm{tr}\Big[\mathbf{T}_i^{\mathsf{T}}C_i\mathbf{T}_i\Big]-e^T\lambda_i\bullet\big(Id^{m\times m}-\mathbf{T}_i^{\mathsf{T}}\mathbf{T}_i\big)\,e\Big]. \tag{10}$$

Since the columns of $\mathbf{T}_i$ are eigenvectors of C_i, and $\Lambda_i=-\lambda_i$, we obtain, by inserting (7) into (10) and making use of (2), the unique maximizer of $\mathcal{L}$

$$\mathcal{L}=\sum_{i=1}^{K}\mathrm{tr}\,\lambda_i \tag{11}$$

given by the m *dominant eigenvalues* of the data-covariance C_i. (It is obvious that the minimizer may not be unique if some eigenvalues around the m-th one (if ordered due to decreasing modulus) are identical. We ignore this pathological situation here).

For the case of $K=1$, we found the well-known theorem which states that the optimal value of the parameter μ is given simply by the mean 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 an assumption about the form of the x_t distribution is needed, i.e., local PCA can be applied for dimension reduction of arbitrary time series.

PCA-based dimension reduction provides a useful tool for data compression. The error of the data-reconstruction from the projected trajectory $y_t=\sum\limits_{i=1}^{K}\gamma_i(t)\mathbf{T}_i^{\mathsf{T}}(x_t-\mu_i)$ is given by the value of the residuum functional (4). This means that the original data set $x(t)$ can be approximated by a reconstructed time series x_t^{red}

$$x_t^{red}=\sum_{i=1}^{K}\gamma_i(t)\,(\mathbf{T}_i y_t+\mu_i) \tag{12}$$

such that the error $\left\| x_t - x_t^{red} \right\|_2^2$ is bounded by the minimum of the residuum functional from (4).

In many interesting cases the standard global PCA–approach (i. e. for $K = 1$) does not result in a meaningful dimension reduction. Let us assume, for example, a time series resulting from a two–dimensional double-well Langevin–dynamics of the form

$$\ddot{x}(t) = -\text{grad}\, V\left(x(t)\right) - \gamma \dot{x}(t) - \sigma \dot{W}, \tag{13}$$

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:

$$\begin{aligned} V(x) = &\sum_{l=1}^{2} a_l \exp\left(-(x - \mu_l^{sys})^{\mathsf{T}} D_l^{sys} (x - \mu_l^{sys})\right) \\ &+6(x - 0.5(\mu_l^{sys} + \mu_l^{sys}))^{\mathsf{T}} (D_1^{sys} + D_2^{sys})(x - 0.5(\mu_l^{sys} + \mu_l^{sys})), \end{aligned} \tag{14}$$

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

The Langevin dynamics in this case produces two clusters of states each associated with the corresponding *metastable* well. The application of PCA 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 PCA with $m = 1$ to each of the clusters separately, we can reduce the value of the residuum–functional 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), indicating 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 "local" dimension reduction by means of PCA. Furthermore, we can ask a more ambitious question: Is it possible to use the local principle dimensions as tokens in the clustering of the time series itself? If possible this will allow to combine clustering of data and dimension reduction in one algorithmic step hopefully leading to synergetic effects and allowing to improve both, clustering of the time series in metastable sets and dimension reduction.

1.1 Direct optimization of the residuum-functional

For a given time series x_t, the minimization of the constrained residuum-functional (4) wrt. the parameters $(\gamma_i(t), \mathbf{T}_i, \mu_i), t = 1, \ldots, T, i = 1, \ldots, K$ can be achieved using an appropriate numerical optimization method. In

order to avoid the problem connected with the high dimension of the resulting optimization problem due to a large number of unknown occupation probabilities $\gamma_i(t)$, we use a wavelet-Galerkin representation of the function $\gamma_i(t)$. We assume that the unknown function $\gamma_i(t)$ can be represented as a finite linear combination of (few) discrete Haar-wavelet functions $\phi(x)$

$$\phi(x) = \xi_{[0,1)} = \begin{cases} 1 & 0 \leq x < 1 \\ 0 & \text{any other} \end{cases} \tag{15}$$

With their help any hidden occupation probability function $\gamma_i(t) \in \mathbf{L}_2(\mathcal{R})$ on any given scale $J \in \mathcal{Z}$ can be represented by a respective scale-specific projection

$$\begin{aligned} P_J\gamma_i(t) &= \sum_{r\in\mathcal{Z}} c_r^i \phi(2^J t - r) \\ c_r^i &= \int_0^1 \gamma_i\left(2^{-J}(r+s)\right) ds \end{aligned} \tag{16}$$

If the number of the ansatz functions in expansion (16) can be assumed to be small, it will allow to project the original high–dimensional optimization problem to the low-dimensional space of the wavelet coefficients c_r^i. The integral transformation between the wavelet representation and the occupation probabilities $\gamma_i(t)$ can be efficiently implemented using the fast Haar-wavelet transformation (FWT) [43].

In our specific implementation of the wavelet-based optimization procedure, we made two simplifying assumptions: (i) we assumed that the occupation probability functions $\gamma_i(t)$ can take only discreet values 0 and 1 (i. e. the occupation probabilities are assumed to be discreet step functions) and (ii) we fixed the upper limit of the Galerkin subspace dimension for each of the optimization runs (i. e., together with the assumption (i) it means that we set the upper limit of transitions between K hidden states). Application of the *adaptive* wavelet approach (where the total number and the scale of the respective wavelet-Galerkin basis functions can be chosen automatically, like it was already demonstrated for other optimization problems [44]) will allow to get rid of those restraining assumptions. This is the matter of ongoing research.

1.2 Hidden Markov Models (HMM)

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. 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, i. e., 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 we assume the state space of hidden states as finite, we thus are concerned with a Markov chain, which is characterized by a 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.

To describe the whole system, the number of hidden states, the transition matrix between them, an initial distribution, and, for each state, a certain probability distribution for the observation are needed .

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

- $S = \{1, 2, ..., K\}$ is a set of K states,
- $V \subset \mathbb{R}^k$ is the observation space,
- $A = (A_{ij})$ is the transition matrix, where $A_{ij} = P(X_{t+1} = j | X_t = i)$ describes the the transition probability from state i to state j,
- ρ_k, $k = 1, \ldots, K$ are probability density functions in the observation space which correspond to the hidden states,
- $\pi = (\pi_i)$ is a stochastic vector, that describes the initial state distribution $\pi_i = P(X_1 = i)$.

Often, the short notation $\lambda = (A, \rho, \pi)$ 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 ρ_k.

1.3 HMM-PCA

Parameter fitting for HMMs can be performed 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 (x_t) as being given and ask for the variation of the probability in terms of the parameters; (X_t) again refers to the sequence of hidden states. The *maximum likelihood principle* simply states that the optimal parameters are given by the absolute maximum of L. Thus, similarly to PCA-based dimension reduction, the maximum likelihood principle is an optimization problem in parameter space.

In order to combine both approaches, we make two assumptions on the observation process: (i) when being in a fixed hidden state, the observed data are distributed according to a multivariate Gaussian distributions ρ_i (where each ρ_i is characterized by a set of corresponding parameters $(\mu_i, \mathbf{T}_i, S_i)$), (ii) the process switching between the hidden states is Markovian, i.e., the

probability of the conformational change depends on the current hidden state only. The first assumption is approximately valid in many cases, e.g., for the first-order differences of logarithms of stock prices (this property of the data can be checked *a posteriori*, see numerical examples below). The second assumption is connected to the characteristic timescale at which the memory of the system under consideration 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

$$P(x_t, X_t|\lambda) = \pi_{X_0} e^{-\frac{1}{2}\left(x_0-\mu_{X_0}\right)^{\mathsf{T}}\mathbf{T}_{X_0}S_{X_0}\mathbf{T}_{X_0}^{\mathsf{T}}\left(x_0-\mu_{X_0}\right)} \prod_{k=1}^{T-1} \frac{A(X_k, X_{k+1})}{\sqrt{(2\pi)^m \det(S_{X_{k+1}})}} \times e^{-\frac{1}{2}\left(x_{k+1}-\mu_{X_{k+1}}\right)^{\mathsf{T}}\mathbf{T}_{X_{k+1}}S_{X_{k+1}}\mathbf{T}_{X_{k+1}}^{\mathsf{T}}\left(x_{k+1}-\mu_{X_{k+1}}\right)}, \quad (17)$$

where A is the transition matrix of the hidden process, $B_i = (\mu_i, \mathbf{T}_i, S_i)$ a set of multivariate Gaussian distribution parameters with $\mu_i \in \mathbb{R}^n$ denoting 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 (17) should be additionally subjected to constraints: (i) the projector orthogonality condition (2), and (ii) 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 optimize the logarithm of the likelihood functional. Writing the resulting log-likelihood functional together with both constraints in Lagrange–form, taking the derivatives wrt. to the model parameters and setting them to zero we get:

$$\mu_i = \frac{\sum_{t=1}^{T} \alpha_i(t)\beta_i(t)x_t}{\sum_{t=1}^{T} \alpha_i(t)\beta_i(t)}, \quad (18)$$

$$\sum_{t=1}^{T} \alpha_i(t)\beta_i(t)(x_t-\mu_i)(x_t-\mu_i)^{\mathsf{T}}\mathbf{T}_i = \mathbf{T}_i S_i, \quad (19)$$

where $\alpha_i(t), \beta_i(t)$ are forward and backward variables (as defined in appendix, see also [22]). 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 (18-19) and those given by localized PCA (6-7) if $\gamma_i(t) = \alpha_i(t)\beta_i(t)$. Only the m dominant eigenvectors of

$$C_i = \sum_{t=1}^{T} \alpha_i(t)\beta_i(t)(x_t-\mu_i)(x_t-\mu_i)^{\mathsf{T}}$$

are needed for the construction of the matrix $\mathbf{T}_i$. Even for large matrices C_i one can compute them efficiently with some iterative subspace method (e.g. Lanczos eigenvalue solver, cf. [45]). In the case of several hidden states

we suggest to use the standard Expectation-Maximization algorithm [46], often also called the Baum-Welch algorithm [16,18]. This approach improves iteratively an initial parameter set, and converges to a local maximum of the likelihood function. Its two steps, the E- and the M-step, are iteratively repeated until the improvement of the likelihood becomes smaller than a given limit. The EM algorithm used herein follows standard procedures; the detailed formula are given in the appendix.

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 K of hidden states, a distribution function for the output of each state, and initial values for all remaining parameters.

2 Numerical Examples

2.1 Langevin Dynamics in 202 Dimensions

As a first example we consider realizations of the Langevin equation (13) with $x = (q, y) \in \mathbb{R}^2 \times \mathbb{R}^N, N = 200$ and the perturbed two–hole potential

$$V(x) = \sum_{l=1}^{2} a_l \exp\left(-(q-\mu_l^{sys})^\mathsf{T} D_l^{sys}(q-\mu_l^{sys})\right) + \frac{1}{2} y^\mathsf{T} D_{bath} y \qquad (20)$$

$$+\delta_0\Big(\cos(2\pi k(x_1+x_2)) + \cos(2\pi k(x_1-x_2))\Big), \qquad (21)$$

where $\delta_0 \ll 1$ is a small perturbation parameter. The N harmonic bath variables are denoted by y, whereas q labels the two "metastable" dimensions that live in the plane of the double well potential. We have chosen the following parameter values: $\mu_1^{sys} = (1.5, 0)^\mathsf{T}$, $\mu_2^{sys} = (0, 1.5)^\mathsf{T}$, $D_1^{sys} = \begin{pmatrix} 0.5 & 0 \\ 0 & 20 \end{pmatrix}$, $D_2^{sys} = \begin{pmatrix} 20 & 0 \\ 0 & 0.5 \end{pmatrix}$, $a_1 = -6$, $a_2 = -6$ such that we get two contiguously placed orthogonal wells and a transition region in between (the contour lines of this potential are shown in the Fig. 4). The parameter matrices D_{bath} and γ have been chosen to be symmetric, positive definite, and tri-diagonal, with 30.0 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 $\Delta t_{Euler} = 0.005$) and total time length 200. Each tenth instance of the resulting time series has been taken for a subsequent parameter estimation (resulting in observation time step $\tau = 0.05$) such that $T = 4.000$. If projected onto the two "metastable" modes q the resulting time series corresponds to a dynamics in a double–well potential with clear transitions between the metastable wells, cf. Fig. 1.

Furthermore, in order to make our model system more realistic and mimic the features inherent in multidimensional systems, we rotate the

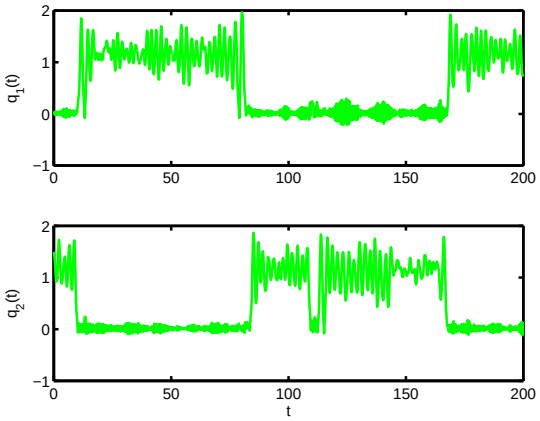

Fig. 1 Projection of the 202-dimensional time series onto the two metastable modes q.

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.

HMM-PCA. Application of the HMM–PCA method indicates the presence of two metastable and one transition state in the time series (see Fig. 2). The phrase "transition state" here means a metastable state with the following two properties: (a) The process under consideration goes through the state whenever it makes a transition between two of the most metastable states of the system, and (b) the transition state is significantly less metastable than the other metastable states.

Clustering of the data with the HMM–PCA–approach in the case of 3 hidden states and reduced dimension $m = 2$ results in the hidden path (also called *Viterbi–path* in the literature) shown in Fig. 3.

The two projectors $\mathbf{T}_i$ are 202×2 matrices whose columns are the eigenvectors of the two largest eigenvalues of the corresponding (local) covariance matrices. In each of the two cases the dominant eigenvalue is significantly larger than the second one; the eigenvectors corresponding to the two dominant eigenvalues are almost identical to the columns of the rotation matrix used to obscure the time series (difference in 2-norm below 0.01 in both cases). This shows that HMM-PCA identified the right subspaces.

In order to interpret the quality of the assignment connected with the Viterbi–path identified by HMM-PCA, we rotate the time series back according to the projectors $\mathbf{T}_i$ identified by HMM-PCA, color the elements according to the Viterbi–path and plot them atop of the original potential surface in (q_1, q_2). As we can see in Fig. 4, 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.

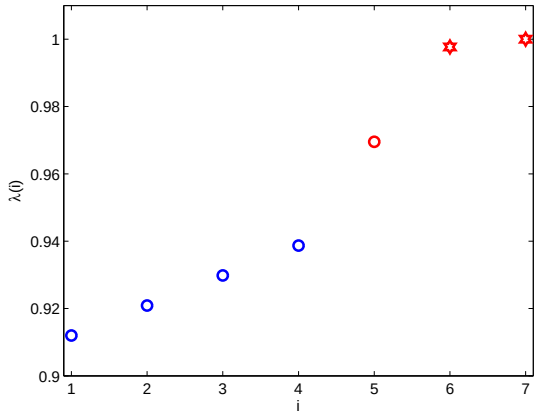

Fig. 2 Spectrum of the transition matrix as calculated for the 202-dimensional time series with HMM–PCA (7 hidden states). It indicates the presence of two metastable states and one transition state (there are two eigenvalues denoted with stars which are located close to 1.0 and one between them and the rest of the spectrum)

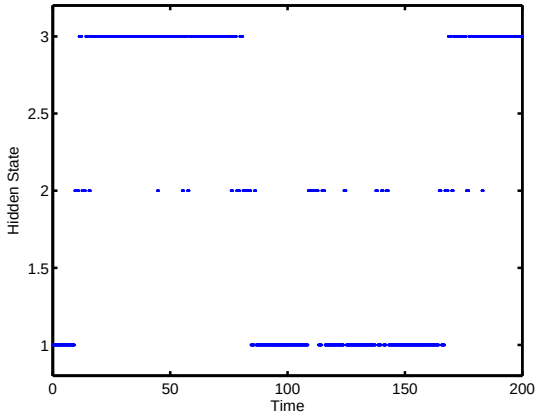

Fig. 3 Viterbi–path resulting from the HMM–PCA clustering of 202–dimensional time series (for 3 hidden states and $m = 2$).

As mentioned before, we hope that dimension reduction provides a tool for data compression in the sense that the original time series can (almost) be reconstructed from the low–dimensional data-set with a help of projectors $\mathbf{T}_i$ and vectors μ_i. The quality of reconstruction in our case is illustrated in Fig. 5. We see that HMM-PCA results in good agreement with the original data. In contrast, the reconstruction based on global PCA (with 2 modes)

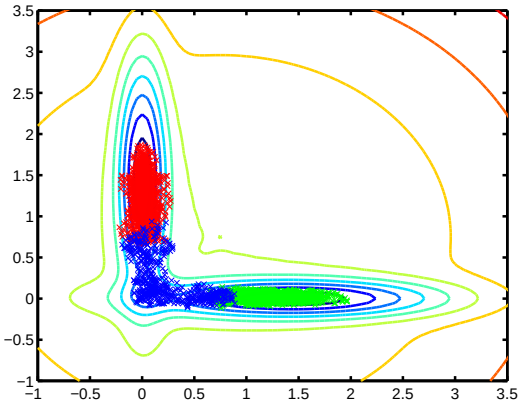

Fig. 4 Coloring of the time series wrt. the Viterbi–path from Fig. 3 (see text). The 202–dimensional data is back–rotated according to the two projectors $\mathbf{T}_i$ and projected onto the two metastable modes (q_1, q_2).

is less efficient and the errors become significant especially at the transition regions.

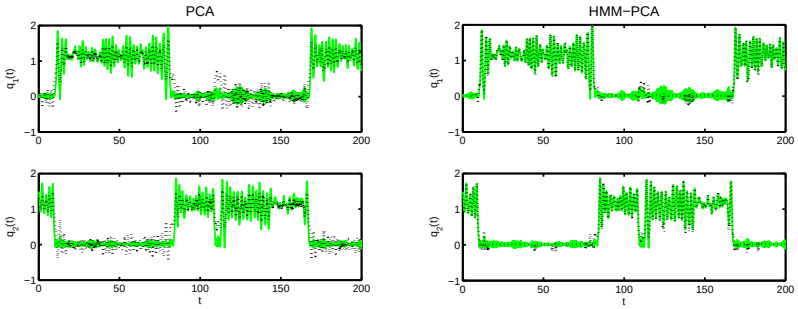

Fig. 5 Original Langevin time series (green) together with time series being reconstructed according to (12) from two–dimensional reduced models based on PCA (left graphics, dotted lines, residuum functional $\mathbf{L} = 2.600$) and HMM-PCA (dotted, right graphics, residuum functional $\mathbf{L} = 1.700$) as described in the text.

2.2 Analysis of the stock market indicators between Jan 2, 1987 and May 12, 2006.

As a second application we analyzed the temporal dynamics of the 12 following stock market indicators: DAX30, Stoxx50(EUR), Stoxx50(USD), NASDAQ Composite, NASDAQ Industrial, NASDAQ Banking, NASDAQ

Insurance, NASDAQ Other Finance, NASDAQ Transportation, NASDAQ Telecommunication, $S\&P500$ and finally Dow Jones Industrial Average (daily closing prices). The data covers the interval between Jan 2, 1987 and May 12, 2006, and thus amounts to 4997 trading days (the data was taken from *www.markt-daten.de*).

As far as the data is obviously non–stationary and has some trend we cannot just assume the multivariate Gaussian distribution to describe the observation probability in the hidden states. However, taking the first differences of logarithms of the time series stationarizes the data [47,23] and makes it possible to apply HMM–PCA.

Application of HMM–PCA with $m = 1$ to the differenced time series indicates the presence of four metastable states (see Fig. 6). The result of the HMM-PCA optimization can be validated by comparison of the resulting Viterbi-path with the plot of the respective probabilities $\gamma_i(t)$ resulting from the direct optimization of the likelihood functional (4). Fig. 7 shows a good agreement of both pathes which validates the applicability of the HMM-PCA-strategy in this case. The differences can be explained by the fact that as it was already explained above, we have implemented a non-adaptive version of the wavelet-based optimization strategy, i. e. in this particular case the maximal number of Haar-wavelet functions was set to 10 for each of the hidden states. Application of the adaptive framework will hopefully increase the similarity between two paths.

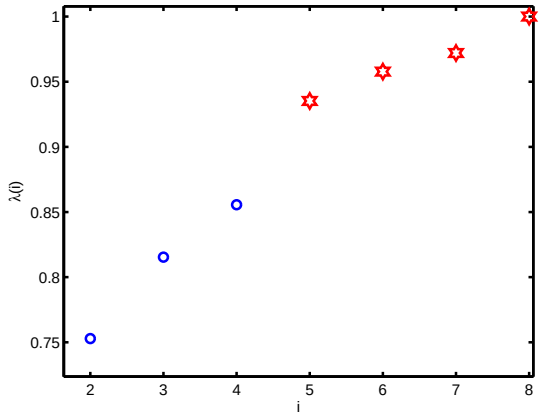

Fig. 6 Spectrum of the transition matrix of the hidden Markov chain (for 7 hidden states, $m = 1$).

The corresponding dominant modes are presented in Fig. 8 and can give an idea about the correlation patterns in the data. This means that positive and negative values of dominant eigenvectors correspond to correlating

and anti–correlating indicators, i.e., whenever the reduced trajectory in the metastable state goes up, the indicators corresponding to positive components of the dominant eigenvector increase and negative ones decrease. As it can be deduced from the comparison of eigenvectors in Fig. 8, the correlation patterns are changing in time. In the first (and less significantly in the third states) the "NASDAQ Banking" indicator is anti–correlated wrt. to all other indicators and in the fourth hidden state it is "NASDAQ Telecommunication".

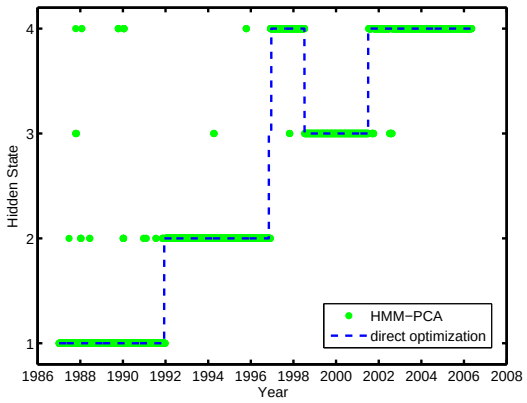

Fig. 7 Comparison of Viterbi–paths resulting from HMM–PCA(dots) and from the direct optimization of the residuum–functional (with the upper number of Haar-wavelet functions set to 10 for each of the hidden states) (4) for 4 hidden states ($m = 1$).

In the same manner as in the previous example, we compare global PCA with HMM–PCA wrt. the data reconstruction quality. As we see from Fig. 10, the one–dimensional projection based on HMM–PCA captures the real dynamics better than standard PCA.

Application of HMM–PCA with $m > 1$. In order to find out whether this type of information (coming only from the one dominant local mode) is enough for qualitative understanding of the correlation behavior in the hidden states, we have to estimate the number m of essential dimensions needed for a reliable representation of the full process in 12–dimensional space of stock indicators. To this end we plot the reconstruction error which describes the l_2–distance between the time–dependent data vector in $\mathbf{R}^{12}$ and its reconstruction from the projection on $\mathbf{R}^m$ for different values of m according to (12). Fig. 9 indicates that, whereas in the first and second hidden states the reconstruction error is actually almost not changing with m, in the third (and somewhat less in the fourth state) the reconstruction error can be brought onto the level of first two states only for $m = 2$. This means

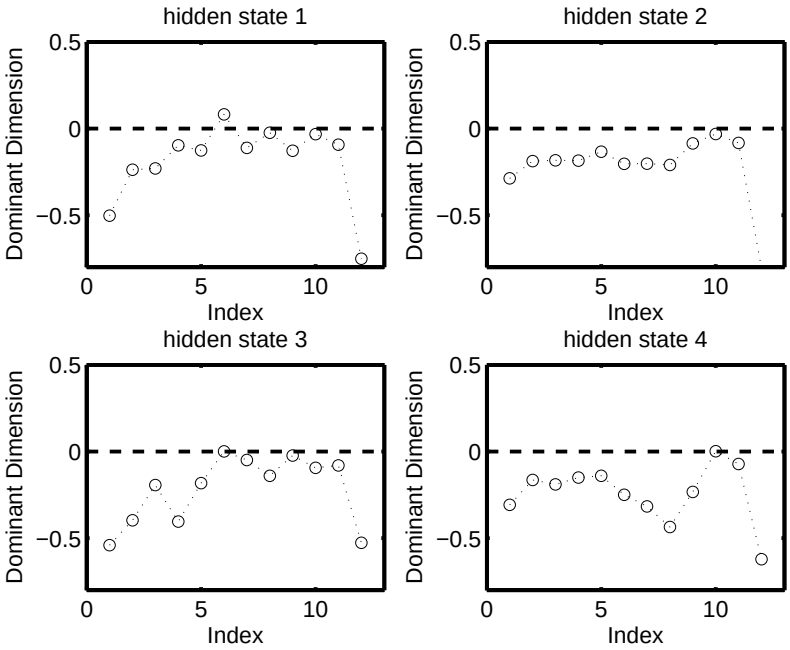

Fig. 8 Dominant dimensions of the optimal projectors $\mathbf{T}_i$ for the four hidden states with the ordering of the stock indexes as being given in the text. Four pictures correspond to the four hidden states in Fig. 7 as being identified via HMM-PCA with $m = 1$. Black dashed lines indicate the zero level.

that in order to get the low-dimensional description of the full process with uniform quality one has to take $m = 1$ for the first and the second hidden states and $m = 2$ for the third and the fourth.

Fig. 11 shows the second dominant modes as resulting from HMM–PCA for $m = 2$ (we get the same Viterbi–path as for $m = 1$). In states 3 and 4 these second modes explain the difference in stock market dynamics around the year 2000. Whereas the "NASDAQ Banking", "NASDAQ Insurance", "NASDAQ Other Finance", "NASDAQ Transportation" and "Dow Jones Industrial" are more or less "flat", other indicators show a "sharp" peak at 2000. As it can be seen from the corresponding elements of second dominant eigenvectors , "flat" indicators are anti-correlating with the "sharp" ones in the second dominant modes.

It should be clear that taking $m > 1$ increases the accuracy of the reproduction in all aspects. However, as we saw one- and two- dimensional local models are already sufficient to capture the essentials of the time series and can help to qualitatively understand the correlation behavior and to figure out the differences between the hidden states. Such kind of analysis can be helpful in the construction of the minimal–risk portfolio [48].

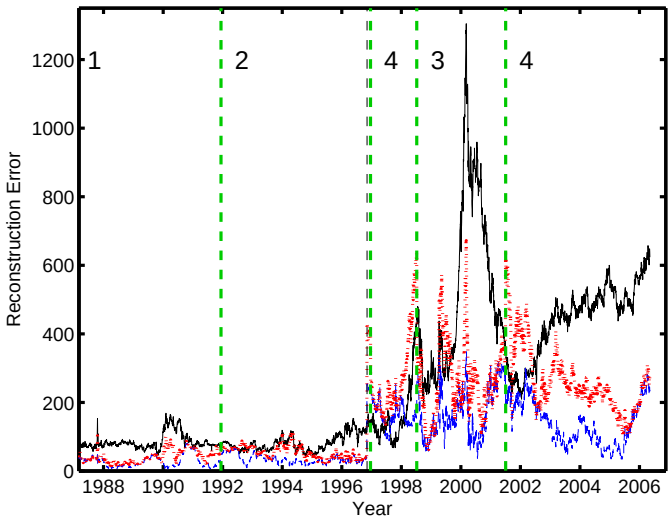

Fig. 9 Comparison of the reconstruction error for the global PCA analysis ($m = 1$, black solid line) with reconstruction errors of HMM-PCA for $m = 1$ (red dotted line) and $m = 2$ (blue dashed line). The green dashed vertical lines and the numbers indicate the switching points of the hidden Viterbi–path from Fig. 7.

2.3 Analysis of some commodities, currencies and stock market indicators between Jan 4, 2000 and April 11, 2006.

As a third application we analyze the temporal dynamics of the 21 daily closing prices of the following stock market titles: Brent oil price in USD, silver price in USD, gold price in USD, Yen price in USD, gold price in Yen, EUR price in USD, gold price in EUR, platinum price in USD, platinum price in EUR, platinum price in Yen, palladium price in USD, palladium price in EUR, DAX 30, NASDAQ Composite, NASDAQ Industrial, NASDAQ Banking, NASDAQ Insurance, NASDAQ Other Finance, NASDAQ Transportation, NASDAQ Telecommunications and $S\&P500$. The data covers the interval between Jan 4, 2000 and April 11, 2006, and thus amounts to 1636 trading days (as in example before, the data was taken from *www.markt-daten.de*).

HMM-PCA analysis ($m = 1$) of the time series of price increments reveals a presence of four hidden states (since there are four dominant eigenvalues in the spectrum of the HMM transition matrix). The result of the HMM-PCA optimization can be validated by comparison of the resulting Viterbi-path with the plot of the respective probabilities $\gamma_i(t)$ resulting from the direct optimization of the likelihood functional (4). As it can be seen from the Fig. 12, the direct optimization of (4) results in $K = 3$ metastable states which can be explained by the fact that the non-Gaussian metastable state 1 in terms of the direct optimization procedure is approximated by

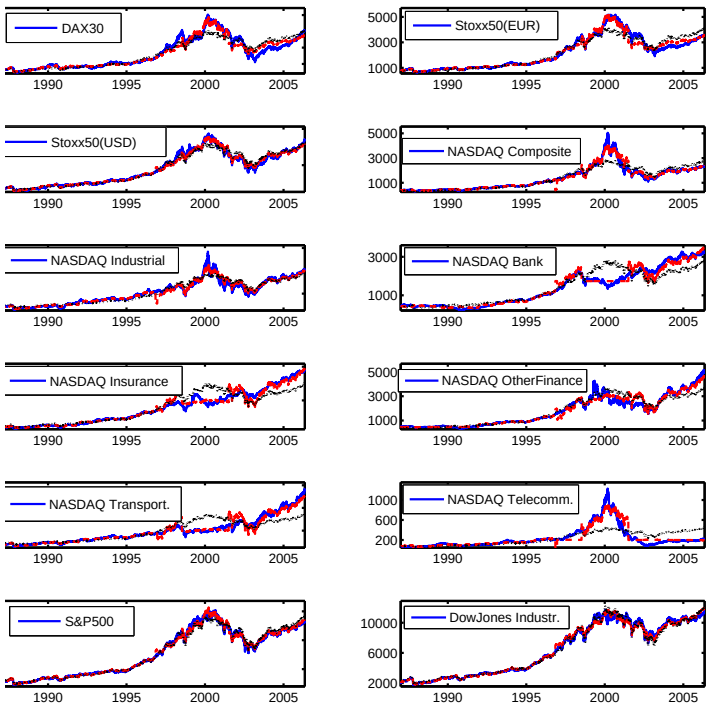

Fig. 10 Original time series (blue) together with their reconstructions from one–dimensional reduced models based on global PCA (black dotted lines, residuum functional $\mathbf{L}_0$) and HMM-PCA (red dashed lines, residuum functional $\mathbf{L} = 0.6\mathbf{L}_0$) both for $m = 1$.

two metastable Gaussian states in context of HMM-PCA. Other identified metastable substates are almost Gaussian (which is verified by applying standard statistical tests to the substate distributions of the data projections on essential dimensions), therefore also the identified Viterbi paths are similar in both cases.

The respective corresponding dominant modes are shown in Figs. 12 and 14. As it can be seen, the most significant entries in the dominant modes are related to just two titles: gold and platinum prices in Yen. In the same manner as in example before, we plot the reconstruction error for different values of m (see Fig. 13) in order to find the adequate number of reduced dimensions for each of the states. It shows up that in order to achieve uniform quality of the dimension reduction, it is enough to take $m = 2$ for the fourth and $m = 1$ for all other states. Moreover, as its can be seen from Fig. 15, the dominant elements of the second dominant mode in the

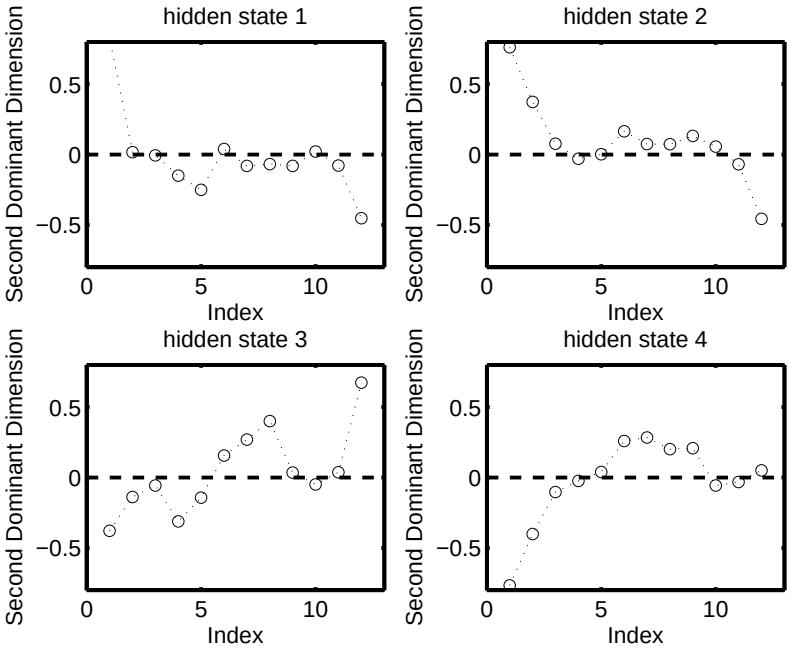

Fig. 11 Second dominant modes of the optimal projectors $\mathbf{T}_i$ for the four hidden states with the ordering of the stock indexes as being given in the text. The four pictures correspond to the four hidden states in Fig. 7. Black dashed lines indicate the zero level.

fourth hidden state are corresponding once again to the gold and platinum prices in Yen. This means that for the overall dimension reduction in the first and second states only the platinum price in Yen is necessary (cf. the upper left and right panels of the Fig. 14) whereas for all other states its enough to take additionally the gold price. The quality of reconstruction of the full dynamics out of these two prices is presented in the Fig. 16. This demonstrates that these two commodities can serve as a reliable indicator of the stock market dynamics.

3 Conclusion

We presented a novel method for simultaneous dimension reduction and identification of metastable phases or regimes from time series data. The problem is formulated in terms of the minimization of the residuum functional describing the distance between the original and the reduced data. We have demonstrated that taking some reasonable assumptions on the underlying process one can apply the HMM–framework to minimize the functional. The resulting numerical method is based on a combination of

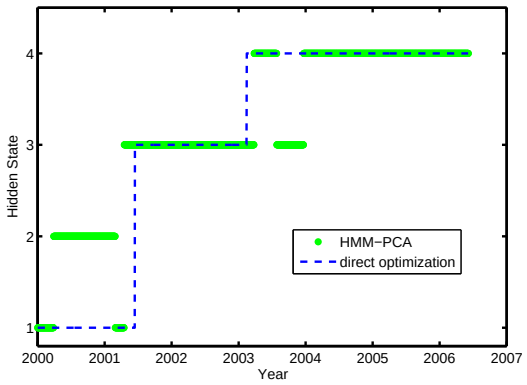

Fig. 12 Viterbi–path resulting from HMM–PCA (for the four hidden states, $m = 1$, dots) and from the direct minimization of the functional (4) ($m = 1$, dashed, the upper number of Haar-wavelet functions set to 10 for each of the hidden states).

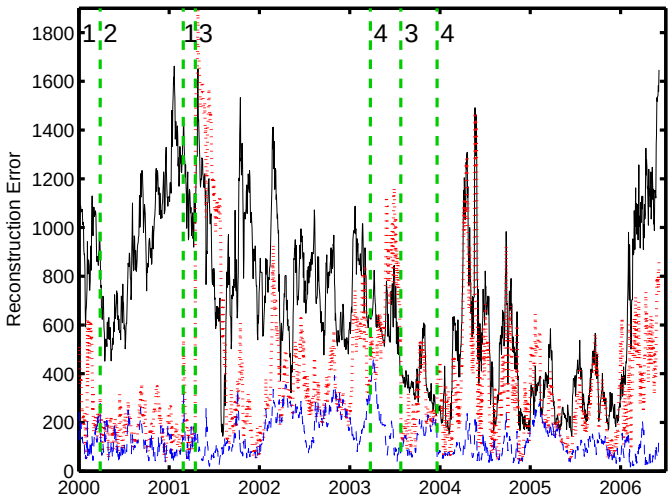

Fig. 13 Comparison of the reconstruction error for the global PCA analysis ($m = 1$, black solid line) with reconstruction errors of HMM-PCA for $m = 1$ (red dotted line) and $m = 2$ (blue dashed line). The green dashed vertical lines and the numbers indicate the switching points of the hidden Viterbi–path from Fig. 12.

the HMM approach and local PCA analysis and has the nice feature that it is linear wrt. the length of data–sequence and so far is suitable for very long time series. Incorporation of local PCA analysis helps to map the clustering problem into low dimensional space. We have demonstrated the application of the method to a model system and to some time series of stock

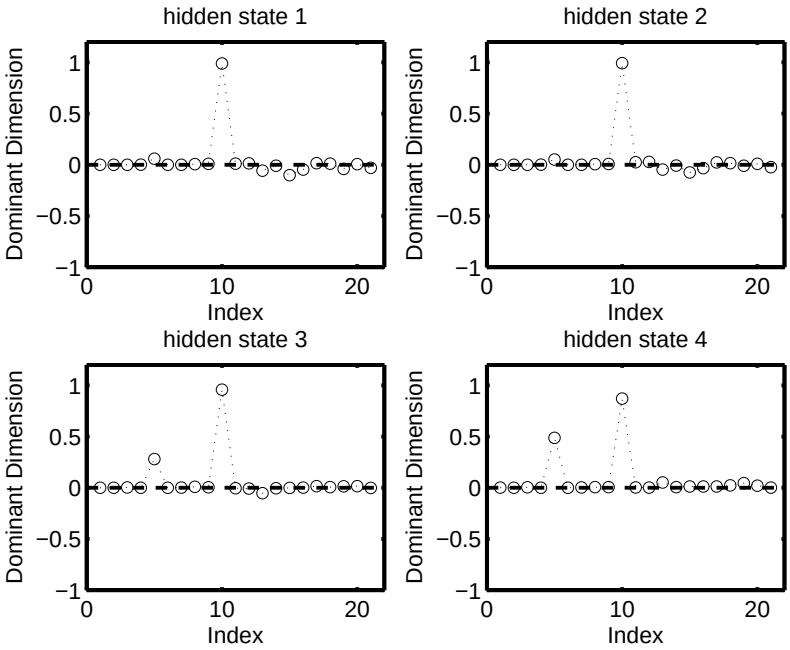

Fig. 14 Dominant modes of the optimal projectors $\mathbf{T}_i$ for the four hidden states with the indexing of the titles as being given in the text. Four pictures correspond to the four hidden states in Fig. 12 as being identified via HMM-PCA with $m = 1$. Black dashed lines indicate the zero level.

market titles and compared the performance of HMM-PCA-method with a direct optimization method based on a discrete wavelet representation of the hidden occupation probabilities. The numerical examples demonstrate the usefulness of the HMM-PCA approach and validate the applicability of the approach to identification of the hidden phases in high–dimensional data. We have shown that in the context of the financial analysis the method can help to find hidden regimes in the high–dimensional data, to extract the leading market indicators and to analyze the principal correlation patterns hidden in the data. We have also demonstrated how the reconstruction error associated with the HMM-PCA representation can help to find the number of essential degrees of freedom in each of the hidden states.

Appendix: EM Steps

There is no known way to analytically determine the model parameters that globally maximize the probability of the given observation sequence. We can, however, estimate λ such that it locally maximizes the likelihood $L(\lambda) = P(x_t \mid \lambda)$. The EM algorithm is a learning algorithm, it iterates two

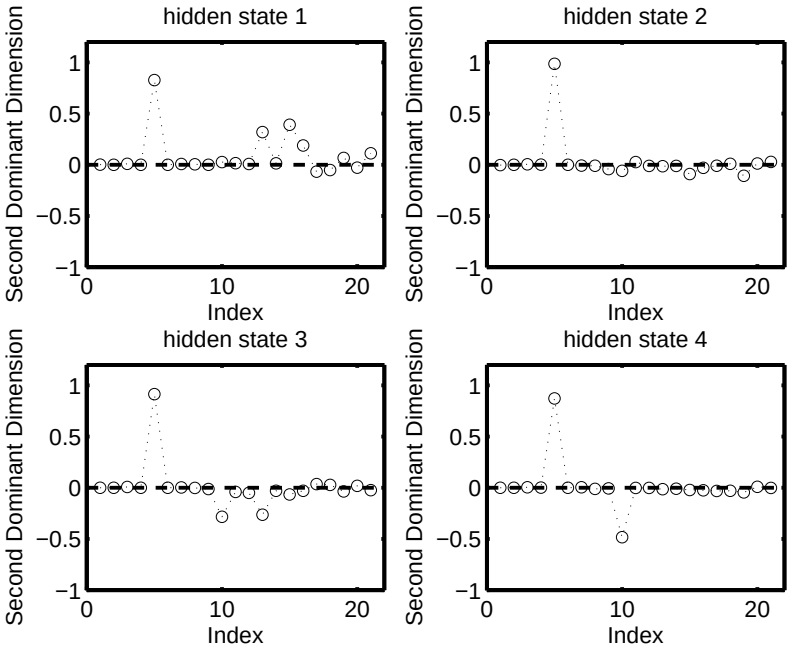

Fig. 15 Second dominant modes of the optimal projectors $\mathbf{T}_i$ for the four hidden states with the ordering of the stock indexes as being given in the text. Four pictures correspond to the four hidden states in Fig. 12. Black dashed lines indicate the zero level.

steps, the Expectation step and the Maximization step. Starting with some initial model λ_0 the steps iteratively refine the model:

- The Expectation-step: In this step the state occupation probability $\gamma_t(i) = P(X_t = i \mid x_t, \lambda)$, and the transition probability $\xi_t(i,j) = P(X_t = i, X_{t+1} = j \mid x_t, \lambda)$, are calculated for each time t in the sequence, given the observation x_t and the current model λ.
- The Maximization-step: This step finds a new model $\hat{\lambda}$ via a set of reestimation formulas. The maximization guarantees that the likelihood does not increase in each single iteration.

In order to calculate the two conditional probabilities of the **E-step** we first define two additional variables

$$\alpha_t(i) = P(x_1 x_2 \ldots x_t, X_t = i \mid \lambda) \tag{22}$$

and

$$\beta_t(i) = P(x_{t+1} x_{t+2} \ldots x_T, X_T = i \mid \lambda), \tag{23}$$

where $\alpha_t(i)$ and $\beta_t(i)$ are forward– and backward–variables respectively. The interpretation of $\alpha_t(i)$ is as follows: it denotes the probability of the

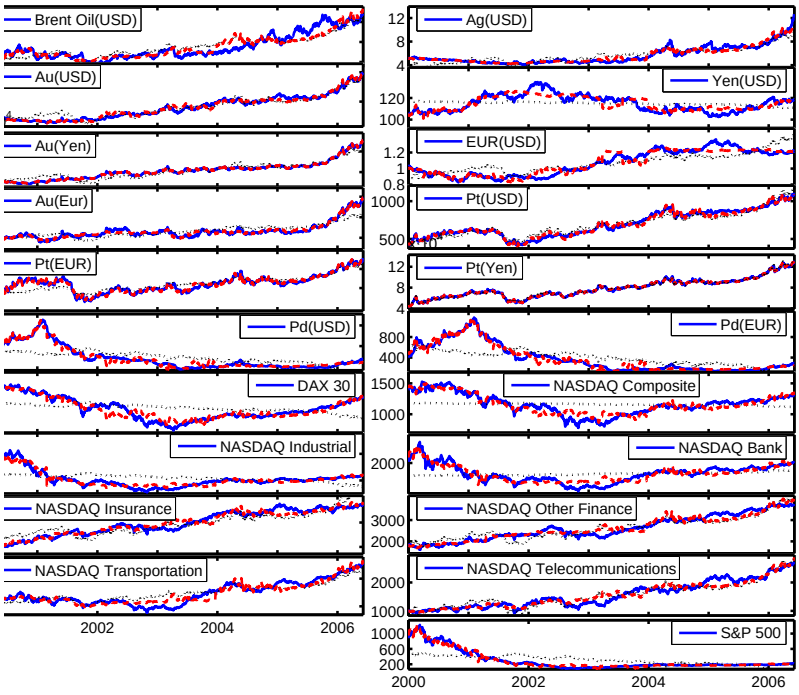

Fig. 16 Original time series (blue) together with their reconstructions from only the gold and platinum prices in Yen (red dashed lines) as compared with the global PCA for $m = 1$ (black dotted lines).

observation sequence up to time t together with the information that the system is in hidden state i at time t conditioned wrt. the given model λ. The following formulas show that the computation of the sequence $\alpha_t(i)$ for the whole sequence is possible with $K^2\,T$ operations:

$$\alpha_1(i) = \pi_i \rho_i(x_1), \qquad 1 \le i \le K$$

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{K} \alpha_t(i) A_{ij}\right] \rho_j(x_{t+1}),$$

$$1 \le t \le T-1,\ 1 \le j \le K.$$

The backward variable $\beta_t(i)$ can be computed with analogous formula:

$$\beta_T(i) = 1, 1 \le i \le K$$

$$\beta_t(i) = \sum_{j=1}^{K} A_{ij} \rho_j(x_{t+1}) \rho_{t+1}(j),$$

$$t = T-1, T-2, \ldots, 1, 1 \le i \le K$$

$$P(x_1 \mid \lambda) = \sum_{i=1}^{K} \beta_1(i) \pi_i \rho_i(x_1) \tag{24}$$

From (22) and (23) one can finally compute the probability for all t as

$$P(x_t \mid \lambda) = \sum_{i=1}^{K} \alpha_t(i)\beta_t(i).$$

The two conditional probabilities of the **E-step** can be calculated efficiently by using the forward-backward variables:

$$\xi_t(i,j) = \frac{\alpha_t(i)A_{ij}\rho_j(x_{t+1})\beta_{t+1}(j)}{P(x_t \mid \lambda)}$$

With these values the probability to be in state i at time t can be expressed as

$$\gamma_t(i) = \sum_{j=1}^{K} \xi_t(i,j).$$

Note that the expected number of visits in the state i to is

$$\sum_{t=1}^{T-1} \gamma_t(i),$$

and the expected number of transitions from i to j is

$$\sum_{t=1}^{T-1} \xi_t(i,j).$$

The **M-step** consists of reestimation formulas for the improved model $\hat{\lambda}$. The estimators for the hidden Markov chain are given by

$$\begin{aligned} \hat{\pi}_i &= \gamma_1(i) \\ \hat{A}_{ij} &= \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \end{aligned} \tag{25}$$

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. A. Amadei, A.B.M. Linssen, and H.J.C Berendsen. Essential dynamics on proteins. *Proteins*, 17:412–425, 1993.
3. D. Frenkel and B Smit. *Understanding Molecular Dynamics: From Algorithms to Applications*. Academic Press, London, 2002.
4. 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.

5. 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.
6. P. Holmes, J.L. Lumley, and G. Berkooz. *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge University Press, 1996.
7. D. Givon, R. Kupferman, and A. Stuart. Extracting macroscopic dynamics: Model problems and algorithms. *Nonlinearity*, 17:R55–R127, 2004.
8. 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.
9. 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.
10. M.A. Balsera, W. Wriggers, Y. Oono, and K. Schulten. Principal Component Analysis and long time protein dynamics. *J. Chem. Phys.*, 100:2567–2572, 1996.
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. J.D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57:357–384, 1989.
18. L.E. Baum. An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. *Inequalities*, 3:1–8, 1972.
19. J.A. Bilmes. *A Gentle Tutorial of the EM Algorithm and its Applications to Parameter Estimation for Gaussian Mixture and Hidden Markov Models. Technical Report*. International Computer Science Institute, Berkeley, 1998.
20. Z. Ghahramani. An introduction to hidden Markov models and Bayesian networks. *Int. J. Pattern Recognition and Artificial Intelligence*, 15(1):9–42, 2001.
21. J. Frydman and P. Lakner. Maximum likelihood estimation of hidden Markov processes. *Ann. Appl. Prob.*, 13(4):1296–1312, 2003.
22. 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.
23. P.J. Brockwell and R.A. Davis. *Introduction to Time Series and Forecasting*. Springer, Berlin, 2002.

24. R.S. Tsay. *Analysis of financial time series.* Wiley Series in Probability and Statistics. 2005.
25. P. Perron and J. Bai. Estimating and testing linear models with multiple structural changes. *Econometrica*, 66:47–78, 1998.
26. A. Harvey and S.J. Koopman. Signal extraction and the formulation of unobserved components models. *The Econometrics Journal*, 3:84–107, 2000.
27. J. Geweke. *The Dynamic Factor Analysis of Economic Time Series.* Latent Variables in Socio-Economic models. North-Holland, 1977.
28. J. Geweke and K. Singleton. Maximum likelihood confirmatory factor analysis of economic time series. *International Economic Review*, 22:37–54, 1981.
29. G.C. Tiao and R.S. Tsay. Model specification in multivariate time series. *Journal of the Royal Statistical Society B*, 51:157–213, 1989.
30. T.W. Anderson. Canonical correlation analysis and reduced rank regression in autoregressive models. *The Annals of Statistics*, 30:1134–1154, 2002.
31. D. Pe na and E.P. Box. Identifying a simplifying structure in time series. *Journal of the American Statistical Association*, 82:836–843, 1987.
32. P.C.M. Molenar, J.G. De Gooijer, and B. Schmitz. Dynamic factor analysis of nonstationary multivariate time sseries. *Psychometrica*, 57(3):333–349, 1992.
33. M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic-factor model: Identification and estimation. *The Review of Economics and Statistics*, 82(4):540–554, 2000.
34. J.H. Stock and M.W. Watson. Macroeconomic forecasting using diffusion indexes. *Journal of Buisiness and Economic Statistics*, 20(2):147–161, 2002.
35. C.R. Nelson and C.-J. Kim. State-space models with regime-switching: Classical and gibbs sampling approaches with applications. *Journal of the American Statistical Association*, 95(452):1373–1374, 2000.
36. H.-M. Krolzig. Markov-switching procedures for dating the euro-zone business cycle. *Vierteljahreshefte zur Wirtschaftsforschung*, 70(3):339–351, 2001.
37. R. Brüggemann. *Model Reduction Methods for Vector Autoregressive Processes.* Lecture Notes in Economics and Mathematical Systems. Springer, Berlin, 2004.
38. S. Kloessner. Empirical evidence: intraday returns are neither symmetric nor Levy processes. In *Proceedings of 2006 South and South East Asia Econometric Society Meeting*, 2006.
39. T. Hediger, A. Passamante, and M.E. Farrell. Characterizing attractors using local intrinsic dimensions calculated by singular value decomposition and information-theoretic criteria. *Phys. Rev. A*, 41:5325, 1990.
40. H. Kantz and T. Schreiber. *Nonlinear Time Series Analysis.* Cambridge University Press, Cambridge, 2004.
41. S.J. Koopman and A. Harvey. Computing observation weigthts for signal extraction and filtering. *Journal of Economic Dynamic Control*, 27:1317–1333, 2003.
42. I.T. Jolliffe. *Principal Component Analysis.* Springer, 2002.
43. G. Strang and T. Nguyen. *Wavelets and Filter Banks.* Wellesley-Cambridge Press, Wellesley, 1997.
44. V. Comincioli, T. Scapolla, G. Naldi, and P. Venini. A wavelet-like galerkin method for numerical solution of variational inequalities arising in elastoplasticity. *Commun. Numer. Meth. Engng*, 16:133–144, 2000.
45. G.H. Golub and C.F. van Loan. *Matrix computations.* The John Hopkins University Press, second edition edition, 1989.

46. 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.
47. G. Box and G. Jenkins. *Time Series Analysis, Forecasting, and Control*. Holden–Day, 1976.
48. S.L. Lee. Correlation shifts and real estate portfolio management. *Journal of Real Estate Portfolio Management*, Jan.–Apr., 2003.