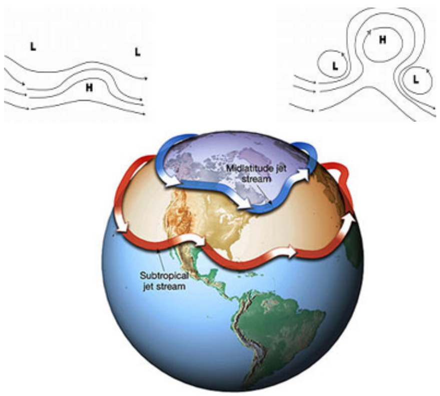
Time Series Analysis in Scientific Computing (24.11.2008-28.11.2008)

I. Horenko, R. Klein, Ch. Schütte

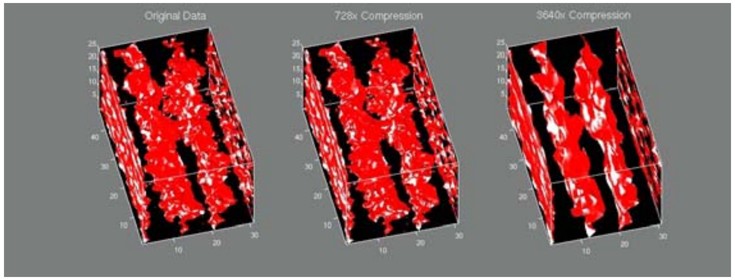
Institute of Mathematics
Freie Universität Berlin (FU)

DFG Research Center ***MATHEON***
„Mathematics in key technologies“

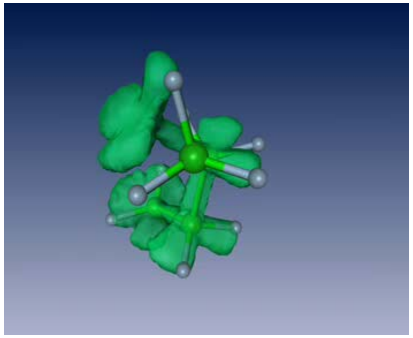
Scientific Computing: Complex Systems



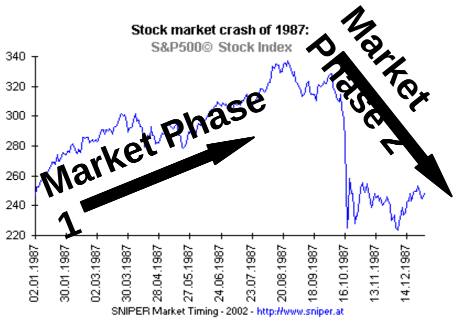
Meteorology/Climate



Fluid Mechanics



Biophysics/Drug Design



Computational Finance

Deterministic description "from first principles"is frequently Unavailable or Unfeasible!

Complex Systems

Properties:

1) *non-stationarity*

2) a lot of *d.o.f.s* are involved (*multidimensionality*)

3) *stochasticity*

4) presence of *hidden phases/regimes*

Aim of the Seminar:

mathematical concepts and methods of multidimensional stochastic time series analysys and identification of hidden phases

Plan of the Seminar (mornings):

Monday (Illia Horenko):

Deterministic view on stochastic processes (direct and inverse numerical problems)

Tuesday (Christof Schütte):

Identification of hidden phases: introduction (K-Means), subspace iteration methods, Expectation-Maximisation algorithm, Gaussian Mixture Models (GMM)

Wednesday (Christof Schütte):

Hidden Markov models (HMM), HMM in multiple dimensions (HMM-VAR)

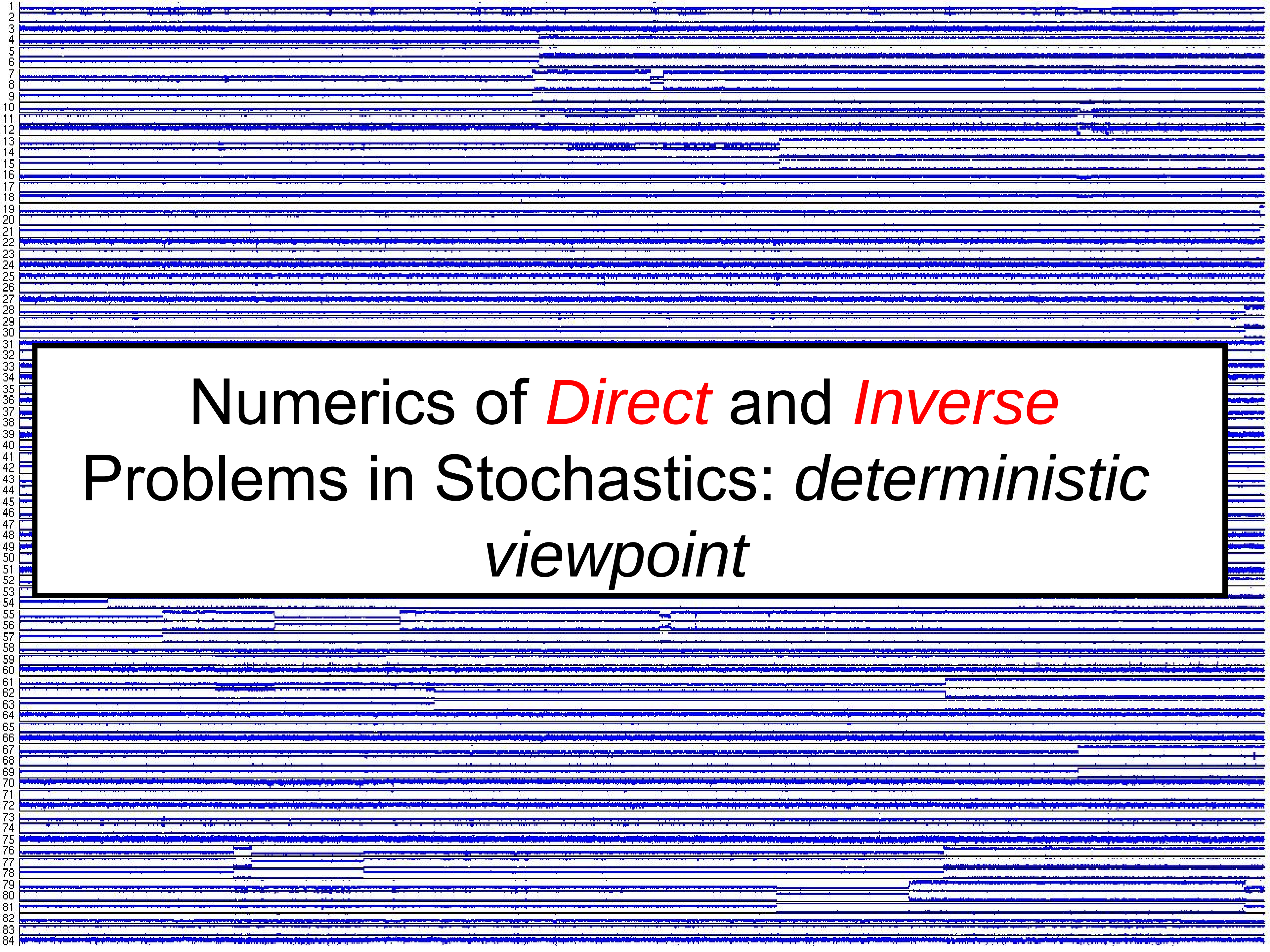
Thursday (Illia Horenko):

Variational approach to time series analysis, finite element methods (FEM) in data analysis (FEM-Clustering)

Friday (Illia Horenko):

Methods of non-stationary time series analysis

Numerics of *Direct* and *Inverse* Problems in Stochastics: *deterministic viewpoint*



Complex Systems

Properties:

1) *non-stationarity*

2) a lot of *d.o.f.s* are involved (*multidimensionality*)

3) *stochasticity*

4) presence of *hidden phases/regimes*

Today we look at:

1) *stochastic processes* and their deterministic interpretation

2) dynamical systems viewpoint

2) inverse problems in stochastics: *functional minimization*

Complex Systems

Properties:

1) *non-stationarity*

2) a lot of *d.o.f.s* are involved (*multidimensionality*)

3) *stochasticity*

4) presence of *hidden phases/regimes*

Today we look at:

1) stochastic processes and their deterministic interpretation

2) dynamical systems viewpoint

2) inverse problems in stochastics: *functional minimization*

Memo I: Probability

A probability space is a measure space such that the measure of the whole space is equal to 1.

In other words: a probability space is a triple $(\Omega, \mathcal{F}, P)$ consisting of a set Ω (called the sample space), a σ-algebra (also called σ-field) $\mathcal{F}$ of subsets of Ω (these subsets are called events), and a measure P on $(\Omega, \mathcal{F})$ such that $P(\Omega) = 1$ (called the probability measure).

Event	Probability
A	$P(A) \in [0, 1]$
not A	$P(A') = 1 - P(A)$
A or B	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ $= P(A) + P(B) \quad \text{if A and B are mutually exclusive}$
A and B	$P(A \cap B) = P(A\|B)P(B)$ $= P(A)P(B) \quad \text{if A and B are independent}$
A given B	$P(A\|B)$

Memo I: Probability

A probability space is a measure space such that the measure of the whole space is equal to 1.

In other words: a probability space is a triple $(\Omega, \mathcal{F}, P)$ consisting of a set Ω (called the sample space), a σ-algebra (also called σ-field) $\mathcal{F}$ of subsets of Ω (these subsets are called events), and a measure P on $(\Omega, \mathcal{F})$ such that $P(\Omega) = 1$ (called the probability measure).

Event	Probability
A	$P(A) \in [0,1]$
not A	$P(A') = 1 - P(A)$
A or B	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ $= P(A) + P(B) \quad$ if A and B are mutually exclusive
A and B	$P(A \cap B) = P(A\|B)P(B)$ $= P(A)P(B) \quad$ if A and B are independent
A given B	$P(A\|B)$

Memo II: Stochastic Process

Given a probability space $(\Omega, \mathcal{F}, P)$, a **stochastic process** (or **random process**) with state space X is a collection of X-valued random variables indexed by a set T ("time"). That is, a stochastic process F is a collection

$$\{F_t : t \in T\}$$

where each F_t is an X-valued random variable.

Probability Density Function:

$$\Pr(4.3 \le X \le 7.8) = \int_{4.3}^{7.8} p(x)\, dx$$

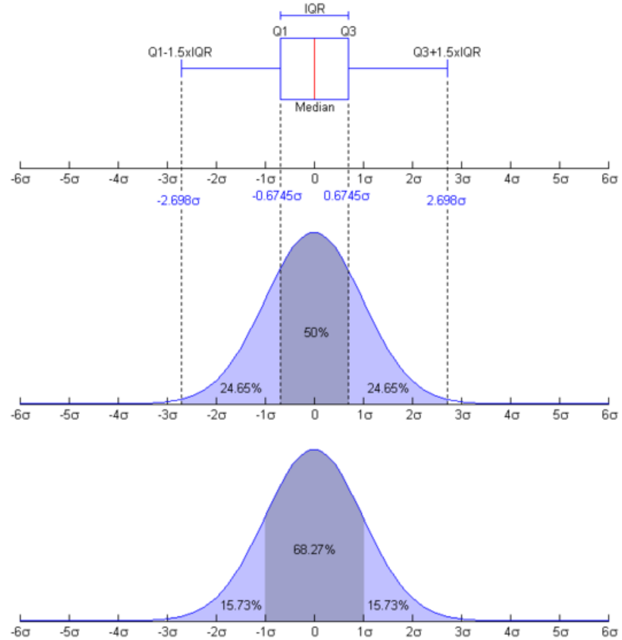
Expectation Value:

$$\mathrm{E}(X) = \int_\Omega X \,\mathrm{d}P \qquad \mu = \int x\, p(x)\, dx$$

Variance:

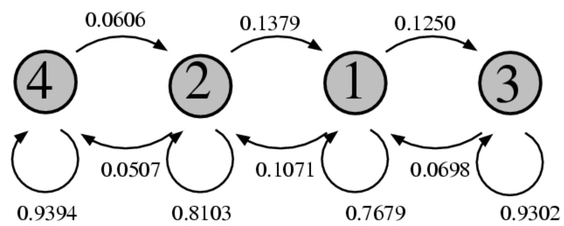
$$\mathrm{Var}(X) = \int (x-\mu)^2\, p(x)\, dx$$

White Noise: $\epsilon_t \sim \mathcal{N}(0,1)$



Classification of Stochastic Process

Discrete State Space,
Discrete Time:
Markov Chain



Discrete State Space,
Continuous Time:
Markov Process

Stochastic Processes

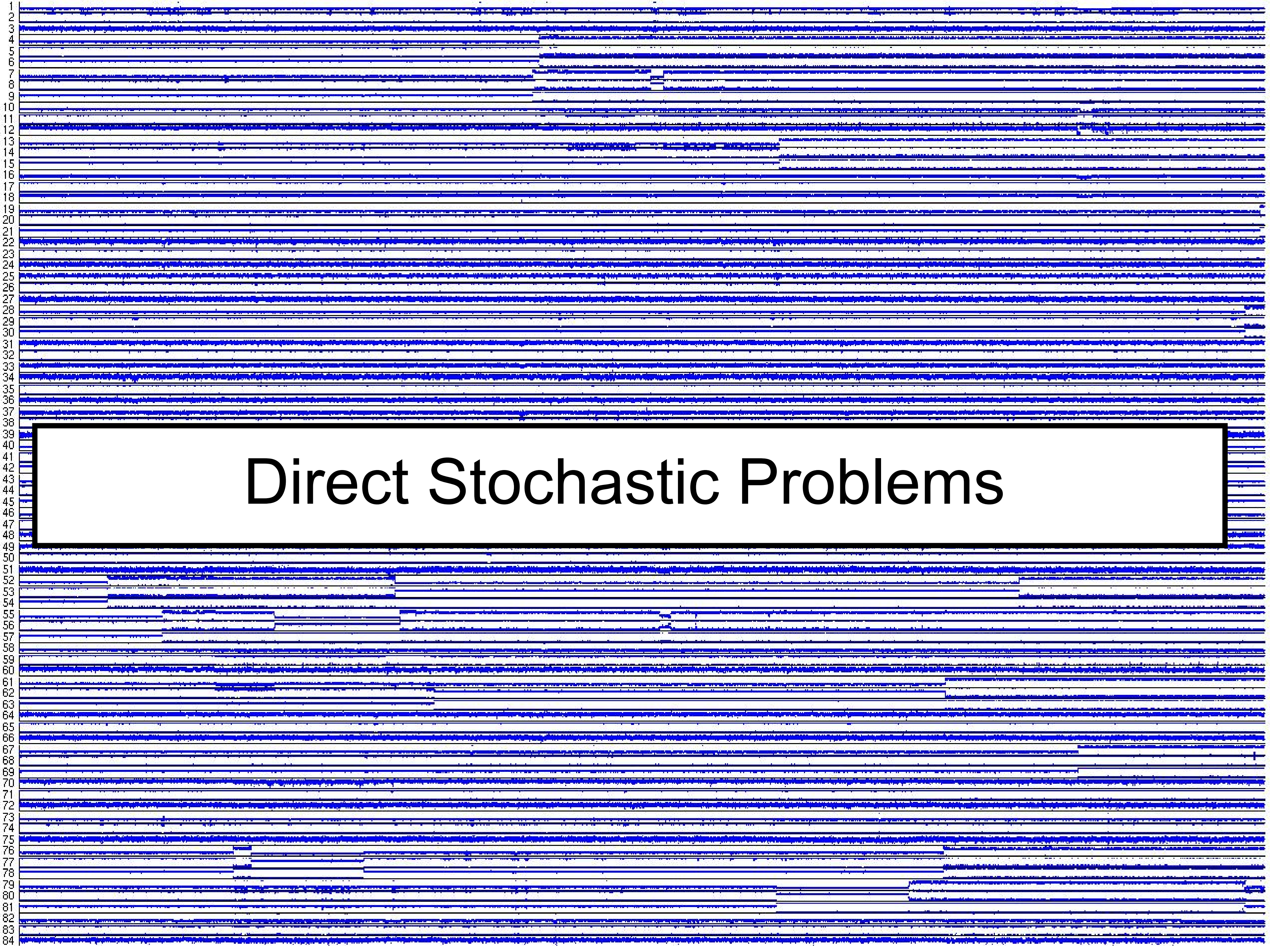
Continuous State Space,
Discrete Time:
Autoregressive Process

$$X_{t+\tau} = \sum_{k=0}^{p} \alpha_k X_{t-k\tau} + \sigma \epsilon_t$$

Continuous State Space,
Continuous Time:
Stochastic Differential Equation

$$dX_t = f\left(X_t, t\right) dt + \sigma dW_t$$

Direct Stochastic Problems



Markov Chains

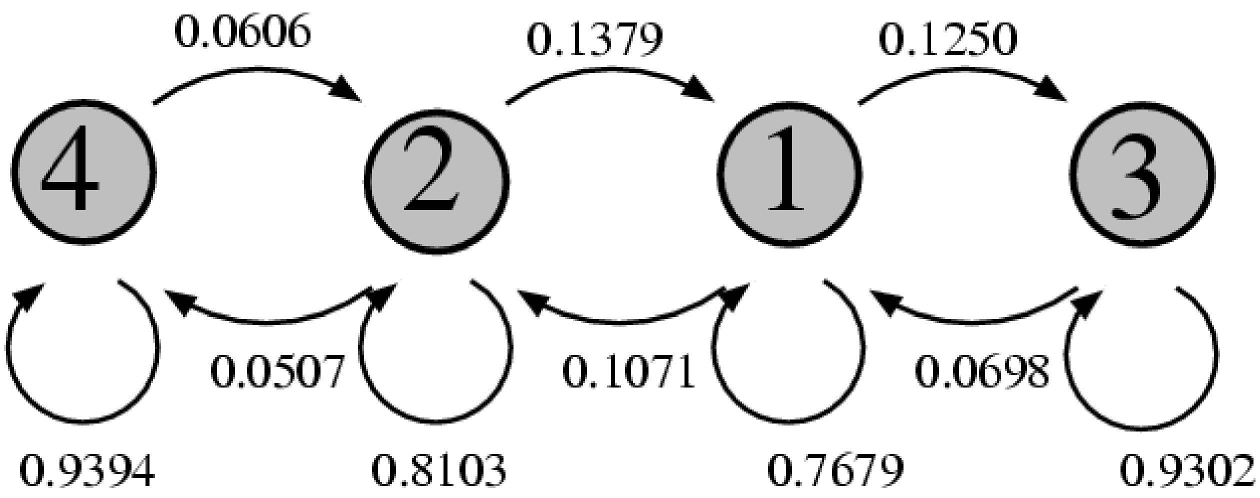
Realizations of the process: $X_t \in s_1, \ldots, s_m$

$$\{X_0, X_\tau, X_{2\tau}, \ldots, \mathbf{X}_{t-\tau}\}$$

Markov-Property:

$$\mathbb{P}\left[X_t = s_j | X_0, X_\tau, X_{2\tau}, \ldots, \mathbf{X}_{t-\tau} = s_j\right] = \mathbb{P}\left[X_t = s_j | \mathbf{X}_{t-\tau} = s_j\right] = P_{ij}\left(t, \tau\right)$$

Example:



Discrete Markov Process…

Realizations of the process: $X_t \in s_1, \ldots, s_m$

$$\{X_0, X_\tau, X_{2\tau}, \ldots, \mathbf{X}_{t-\tau}\}$$

Markov-Property:

$$\mathbb{P}\left[X_t = s_j | X_0, X_\tau, X_{2\tau}, \ldots, \mathbf{X}_{t-\tau} = s_j\right] = \mathbb{P}\left[X_t = s_j | \mathbf{X}_{t-\tau} = s_j\right] = P_{ij}\left(t, \tau\right)$$

State Probabilities: $\pi(t) = (\pi_1(t), \pi_2(t), \ldots, \pi_m(t))$ $\boxed{\pi_i(t) = \mathbb{P}[X_t = s_i]}$

$$\pi_i(t+\tau) = \sum_{k=1}^{m} \mathbb{P}[X_{t+\tau} = s_i, X_t = s_k]$$

Discrete Markov Process…

Realizations of the process: $X_t \in s_1, \ldots, s_m$

$$\{X_0, X_\tau, X_{2\tau}, \ldots, \mathbf{X}_{t-\tau}\}$$

Markov-Property:

$$\mathbb{P}\left[X_t = s_j | X_0, X_\tau, X_{2\tau}, \ldots, \mathbf{X}_{t-\tau} = s_j\right] = \mathbb{P}\left[X_t = s_j | \mathbf{X}_{t-\tau} = s_j\right] = P_{ij}\left(t, \tau\right)$$

State Probabilities: $\pi(t) = (\pi_1(t), \pi_2(t), \ldots, \pi_m(t))$ $\boxed{\pi_i(t) = \mathbb{P}[X_t = s_i]}$

$$\pi_i(t+\tau) = \sum_{k=1}^{m} \mathbb{P}[X_{t+\tau} = s_i, X_t = s_k]$$

$$= \sum_{k=1} P_{ki}(t)\pi_k(t)$$

Discrete Markov Process…

Realizations of the process: $X_t \in s_1, \ldots, s_m$

$$\{X_0, X_\tau, X_{2\tau}, \ldots, \mathbf{X}_{t-\tau}\}$$

Markov-Property:

$$\mathbb{P}\left[X_t = s_j | X_0, X_\tau, X_{2\tau}, \ldots, \mathbf{X}_{t-\tau} = s_j\right] = \mathbb{P}\left[X_t = s_j | \mathbf{X}_{t-\tau} = s_j\right] = P_{ij}\left(t, \tau\right)$$

State Probabilities: $\pi(t) = (\pi_1(t), \pi_2(t), \ldots, \pi_m(t))$ $\boxed{\pi_i(t) = \mathbb{P}[X_t = s_i]}$

$$\pi_i(t+\tau) = \sum_{k=1}^{m} \mathbb{P}[X_{t+\tau} = s_i, X_t = s_k]$$

Discrete Markov Process...

Realizations of the process: $X_t \in s_1, \ldots, s_m$

$$\{X_0, X_\tau, X_{2\tau}, \ldots, \mathbf{X}_{t-\tau}\}$$

Markov-Property:

$$\mathbb{P}\left[X_t = s_j | X_0, X_\tau, X_{2\tau}, \ldots, \mathbf{X}_{t-\tau} = s_j\right] = \mathbb{P}\left[X_t = s_j | \mathbf{X}_{t-\tau} = s_j\right] = P_{ij}\left(t, \tau\right)$$

State Probabilities: $\pi(t) = (\pi_1(t), \pi_2(t), \ldots, \pi_m(t))$ $\boxed{\pi_i(t) = \mathbb{P}[X_t = s_i]}$

$$\pi_i(t+\tau) = \sum_{k=1}^{m} \mathbb{P}[X_{t+\tau} = s_i, X_t = s_k]$$

$$= \sum_{k=1} P_{ki}(t)\pi_k(t)$$

Discrete Markov Process…

Realizations of the process: $X_t \in s_1, \ldots, s_m$

$$\{X_0, X_\tau, X_{2\tau}, \ldots, \mathbf{X}_{t-\tau}\}$$

Markov-Property:

$$\mathbb{P}\left[X_t = s_j | X_0, X_\tau, X_{2\tau}, \ldots, \mathbf{X}_{t-\tau} = s_j\right] = \mathbb{P}\left[X_t = s_j | \mathbf{X}_{t-\tau} = s_j\right] = P_{ij}\left(t, \tau\right)$$

State Probabilities: $\pi(t) = (\pi_1(t), \pi_2(t), \ldots, \pi_m(t))$ $\boxed{\pi_i(t) = \mathbb{P}[X_t = s_i]}$

$$\pi_i(t+\tau) = \sum_{k=1}^{m} \mathbb{P}[X_{t+\tau} = s_i, X_t = s_k]$$

$$= \sum_{k=1} P_{ki}(t)\pi_k(t)$$

$$\boxed{\pi(t+\tau) = \pi(t)P(t,\tau)}$$

This Equation is Deterministic

Continuous Markov Process…

$$\pi(t+\tau) = \pi(t)P(t,\tau)$$

Continuous Markov Process…

$$\pi(t+\tau) = \pi(t)P(t,\tau)$$

$$\frac{\pi(t+\tau) - \pi(t)}{\tau} = \frac{\pi(t)P(t,\tau) - \pi(t)}{\tau}$$

Continuous Markov Process…

$$\pi(t+\tau) = \pi(t)P(t,\tau)$$

$$\frac{\pi(t+\tau) - \pi(t)}{\tau} = \frac{\pi(t)P(t,\tau) - \pi(t)}{\tau}$$

Continuous Markov Process…

$$\pi(t+\tau) = \pi(t)P(t,\tau)$$

$$\frac{\pi(t+\tau)-\pi(t)}{\tau} = \frac{\pi(t)P(t,\tau)-\pi(t)}{\tau}$$

$$\lim_{\tau\to 0}\frac{\pi(t+\tau)-\pi(t)}{\tau} = \lim_{\tau\to 0}\frac{\pi(t)P(t,\tau)-\pi(t)}{\tau}$$

$$\mathcal{G}(t) = \lim_{\tau\to 0}\frac{P(t,\tau)-\mathcal{I}}{\tau}$$

infenitisimal

generator

Continuous Markov Process…

$$\pi(t+\tau) = \pi(t)P(t,\tau)$$

$$\frac{\pi(t+\tau)-\pi(t)}{\tau} = \frac{\pi(t)P(t,\tau)-\pi(t)}{\tau}$$

$$\lim_{\tau\to 0}\frac{\pi(t+\tau)-\pi(t)}{\tau} = \lim_{\tau\to 0}\frac{\pi(t)P(t,\tau)-\pi(t)}{\tau}$$

$$\mathcal{G}(t) = \lim_{\tau\to 0}\frac{P(t,\tau)-\mathcal{I}}{\tau}$$

infenitisimal

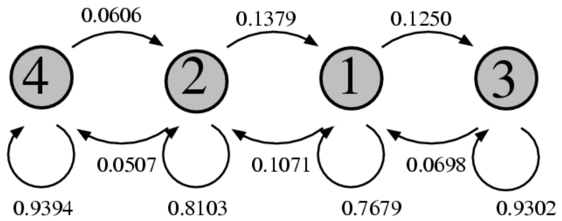
generator

$$\dot{\pi} = \pi\mathcal{G}(t)$$

This Equation is Deterministic: ODE

Continuous State Space

Discrete State Space,
Discrete Time:
Markov Chain



Discrete State Space,
Continuous Time:
Markov Process

Stochastic Processes

Continuous State Space,
Discrete Time:
AR(1)

Continuous State Space,
Continuous Time:
Stochastic Differential Equation

$$X_{t+\tau} = X_t + \tau f(X_t) + \sigma\sqrt{\tau}\epsilon_t$$

$$dX_t = f(X_t, t)\,dt + \sigma dW_t$$

Markov Process in **R**

Realizations of the process: $X_t \in \mathbb{R} \quad \{X_0, X_\tau, X_{2\tau}, \ldots, \mathbf{X}_{t-\tau}\}$

Process:

$$X_{t+\tau} = X_t + \tau f(X_t) + \sigma\sqrt{\tau}\epsilon_t$$

$$\pi(X, t+\tau) = \lim_{\epsilon \longrightarrow 0} \mathbb{P}[X_t \in (X-\epsilon, X+\epsilon)]$$

Markov Process in **R**

Realizations of the process: $X_t \in \mathbb{R}$ $\{X_0, X_\tau, X_{2\tau}, \ldots, \mathbf{X}_{t-\tau}\}$

Process:

$$X_{t+\tau} = X_t + \tau f(X_t) + \sigma\sqrt{\tau}\epsilon_t$$
$$\pi(X, t+\tau) = \lim_{\epsilon \longrightarrow 0} \mathbb{P}[X_t \in (X-\epsilon, X+\epsilon)]$$

If $\pi(X,t) = \delta\left(X - X_t\right)$ *then* $\pi(X, t+\tau) = \mathcal{N}(X_t + \tau f(X_t), \tau\sigma^2)$

Markov Process in **R**

Realizations of the process: $X_t \in \mathbb{R}$ $\{X_0, X_\tau, X_{2\tau}, \ldots, \mathbf{X}_{t-\tau}\}$

Process:

$$X_{t+\tau} = X_t + \tau f(X_t) + \sigma\sqrt{\tau}\epsilon_t$$
$$\pi(X, t+\tau) = \lim_{\epsilon \longrightarrow 0} \mathbb{P}[X_t \in (X-\epsilon, X+\epsilon)]$$

If $\pi(X,t) = \delta(X - X_t)$ *then* $\pi(X, t+\tau) = \mathcal{N}(X_t + \tau f(X_t), \tau\sigma^2)$

Representing the arbitrary intial probability density $\pi(X,t)$ *with Dirac-testfunctions* $\delta(X - X_t)$ *results in:*

$\tau \longrightarrow 0$

$$\pi(X, t+\tau) = \exp\{\frac{\tau\sigma^2}{2}\partial_X^2 + \tau f(X)\partial_X\}\pi(X,t)$$

(follows also from Ito's formula)

This Equation is Deterministic

Markov Process in **R**

Infenitisimal *Generator:*

$$\mathbf{G} = \lim_{\tau \to 0} \frac{\exp\{\frac{\tau\sigma^2}{2}\partial_X^2 + \tau f(X)\partial_X\} - \mathbf{I}}{\tau}$$

$$= \frac{\sigma^2}{2}\partial_X^2 + f(X)\partial_X$$

Markov Process Dynamics:

$$\boxed{\partial_t \pi(X,t) = \mathbf{G}\pi(X,t)}$$

This Equation is Deterministic: PDE

Numerics of Stochastic Processes

$$\pi(t+\tau) = \pi(t)P(t,\tau)$$

$\tau \to 0$

$$\dot{\pi} = \pi\mathcal{G}(t)$$

Discrete State Space,
Discrete Time:
Markov Chain

Discrete State Space,
Continuous Time:
Markov Process

Continuous State Space,
Discrete Time:
Autoregressive Process

Continuous State Space,
Continuous Time:
Stochastic Differential Equation

$\tau \to 0$

$$\pi(X,t+\tau) = \exp\{\frac{\tau\sigma^2}{2}\partial_X^2 + \tau f(X)\partial_X\}\pi(X,t)$$

$$\partial_t\pi(X,t) = \mathbf{G}\pi(X,t)$$

Numerics of Stochastic Processes

$$\pi(t+\tau) = \pi(t)P(t,\tau)$$

Discrete State Space,
Discrete Time:
Markov Chain

$\tau \to 0$

$$\dot{\pi} = \pi\mathcal{G}(t)$$

Discrete State Space,
Continuous Time:
Markov Process

Deterministic Numerics of ODEs and PDEs!

Continuous State Space,
Discrete Time:
Autoregressive Process

$$\pi(X,t+\tau) = \exp\{\frac{\tau\sigma^2}{2}\partial_X^2 + \tau f(X)\partial_X\}\pi(X,t)$$

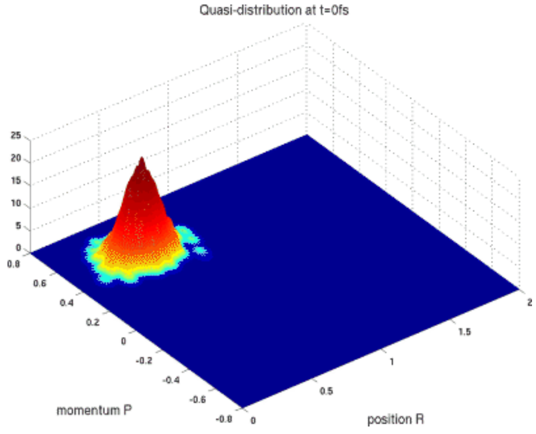
$\tau \to 0$

Continuous State Space,
Continuous Time:
Stochastic Differential Equation

$$\partial_t \pi(X,t) = \mathbf{G}\pi(X,t)$$

Intermediate Conclusions…

1) *Numerical Methods from ODEs and (multidimensional) PDEs like Runge-Kutta-Methods, FEM and (adaptive) Rothe particle methods are applicable to stochatic processes*

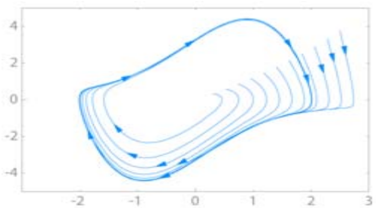


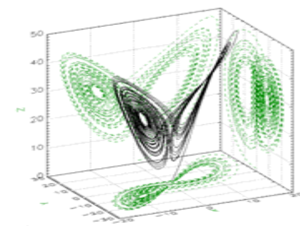
<u>Adaptive PDE particle methods in multiple dimensions:</u>
H./Weiser, JCC **24***(15), 2003*
H./Weiser/Schmidt/Schütte, JCP **120***(19), 2004*
H./Lorenz/Schütte/Huisinga, JCC **26***(9), 2005*
Weiße/H./Huisinga, LNCS **4216***, 2006*

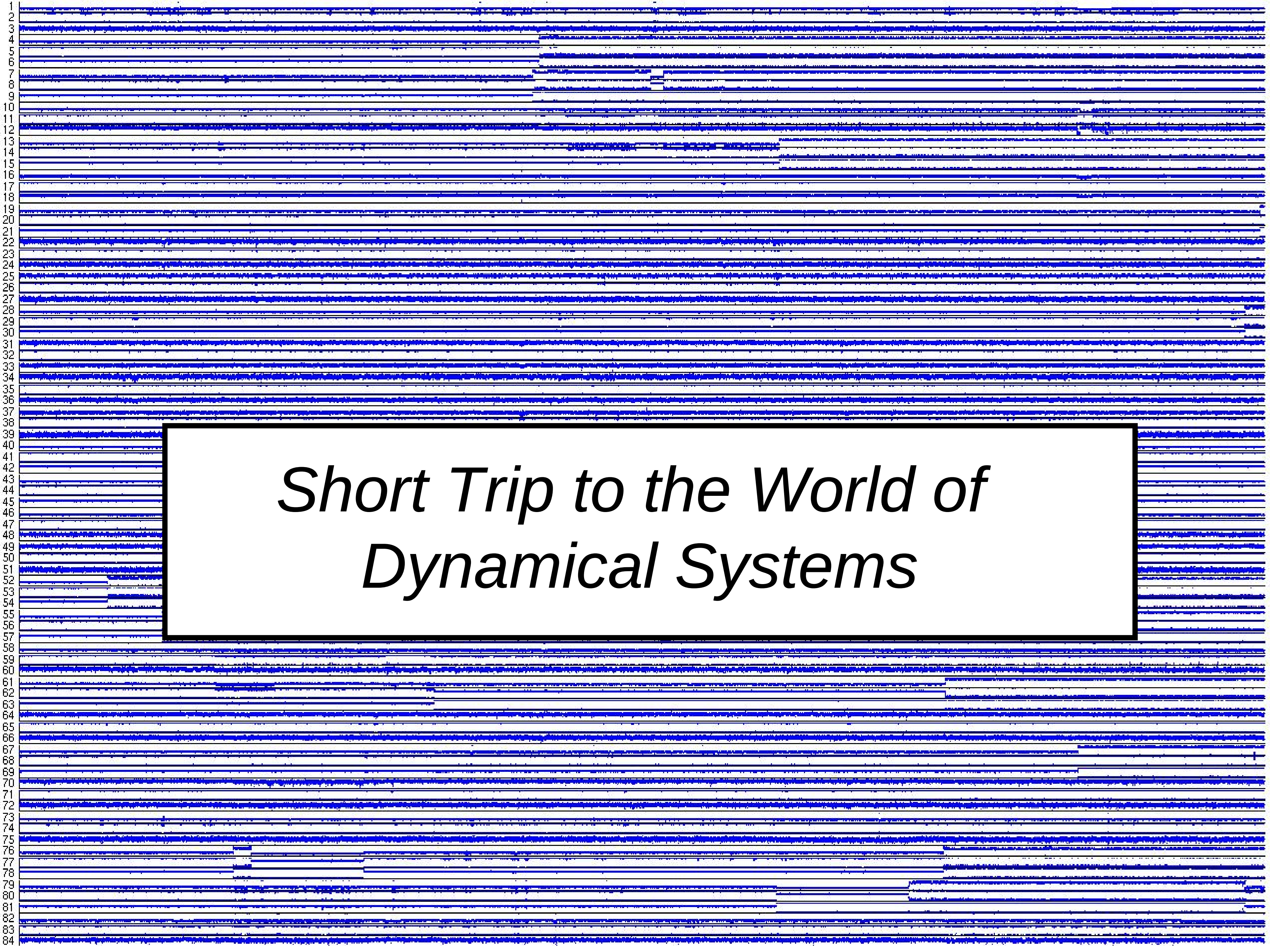
2) *Monte-Carlo-Sampling of resulting p.d.f.'s*

Stochastic Numerics = ODE/PDE numerics + Rand. Numb. Generator

3) *Concepts from the Theory of Dynamical Systems are Applicable (Whitney and Takens theorems, model reduction by identification of attractors)*







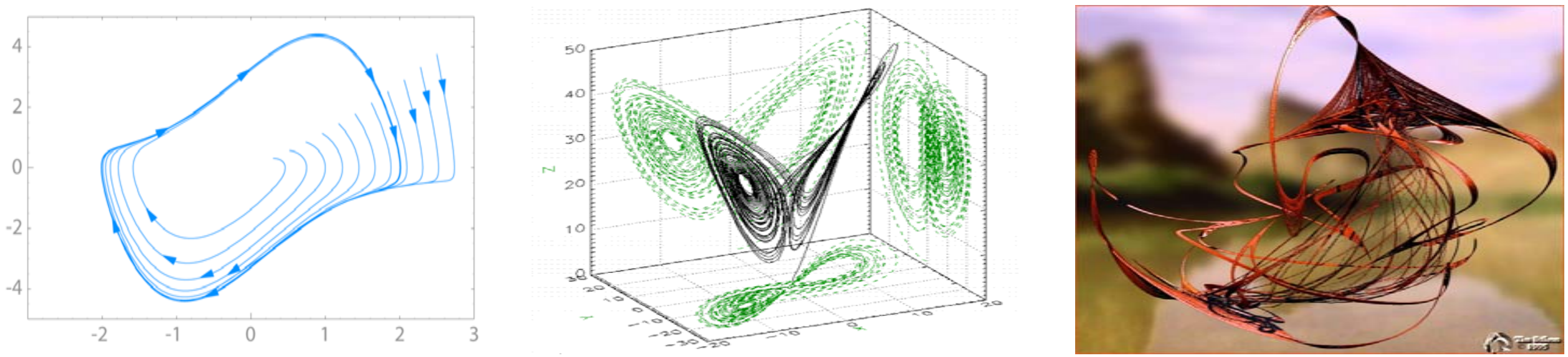
Short Trip to the World of Dynamical Systems

Dynamical systems viewpoint

Essential Dynamics of many dynamical systems is defined by *attractors*.

Attractor $\mathbf{A}$ stays invariant under the *flow operator* f_τ

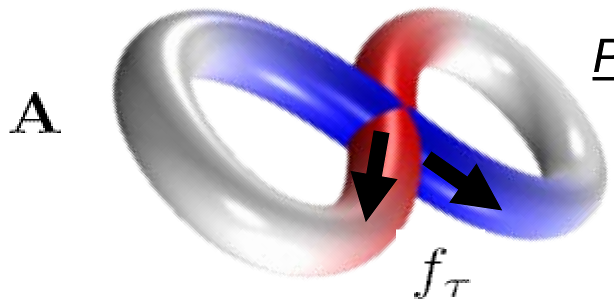
Figures from http:\\www.wikipedia.org



Separation of informative (attractor) and non-informative (rest) parts of the space leads to dimension reduction

How to localize the attractor?

Attractors can have very complex, even fractal geometry



Problem: Euclidean distance cannot be used as a measure for the relative neighbourhood of attractor elements.

Strategy: the data have to be "embedded" into Eucllidean space

Whitney embedding theorem (Whitney, 1936) : sufficiently smooth connected m-dimensional manifolds can be smoothly embedded in $(2m+1)$-dimensional Euclidean space.

How to construct the embedding?

Takens' Embedding *(Takens, 1981)*

$z_t \in \mathbf{M} \subset \mathbf{R}^d$ *and* $f_\tau : \mathbf{M} \rightarrow \mathbf{M}$ *is a smooth map.*

Assume that f_τ *has an attractor* $\mathbf{A}$ *with dimension*

$m << d$. *Let* $\alpha : \mathbf{A} \rightarrow \mathbf{R}^1$, $\alpha \in \mathcal{C}^2$ *be a "proper"*

measurement process. Then

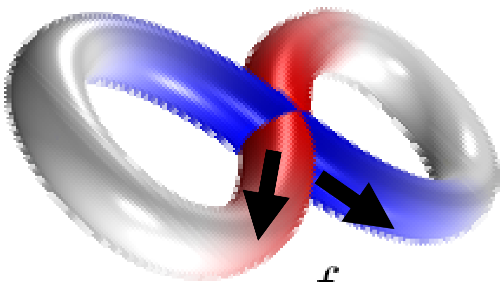
$$\phi_\alpha(z) = (\alpha(z), \alpha(f_\tau(z)), \ldots, \alpha(f_{2m\tau}(z)))$$

is a $(2m+1)$ *-dimensional embedding in Euclidean space*

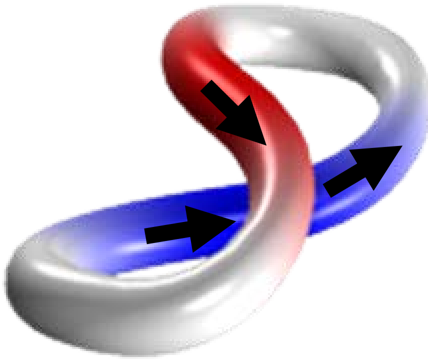
Illustration of Takens' Embedding

The image of $\phi_\alpha(z)$ is unfolded in $\mathbf{R}^{2m+1}$

$\mathbf{A}$



f_τ



Problems:

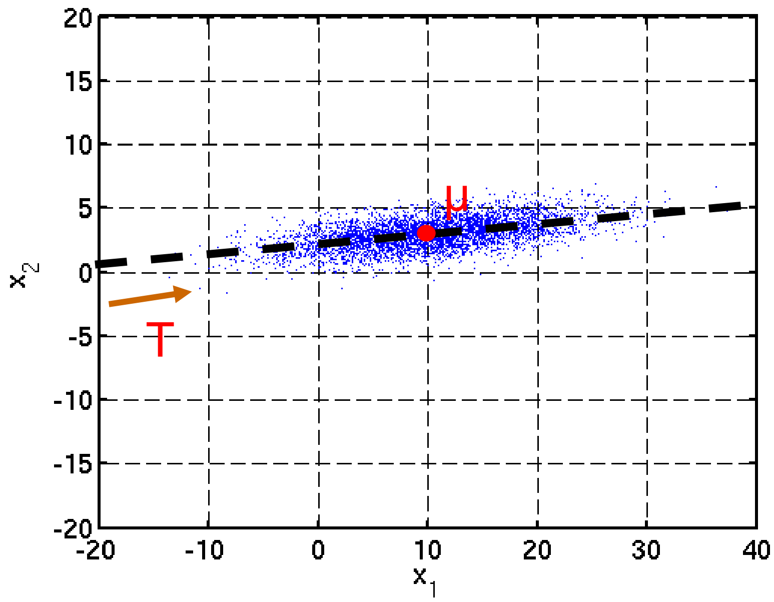
- how to "filter out" the attractive subspace?
- how to connect the changes in attractive subspace with hidden phase?

Strategy:

Let $\{z_t\}_{t=1,\ldots,T}, z_t \in \mathbf{R}^d$. Define a new variable $\{x_t^q\} = \{z_t, z_{t-1}, \ldots, z_{t-q}\}$. Let the attractor $\mathbf{A}$ for each of the hidden phases be contained in a distinct *linear manifold* defined via $\mathbf{T} \in \mathbf{R}^{nxm}, m << d$

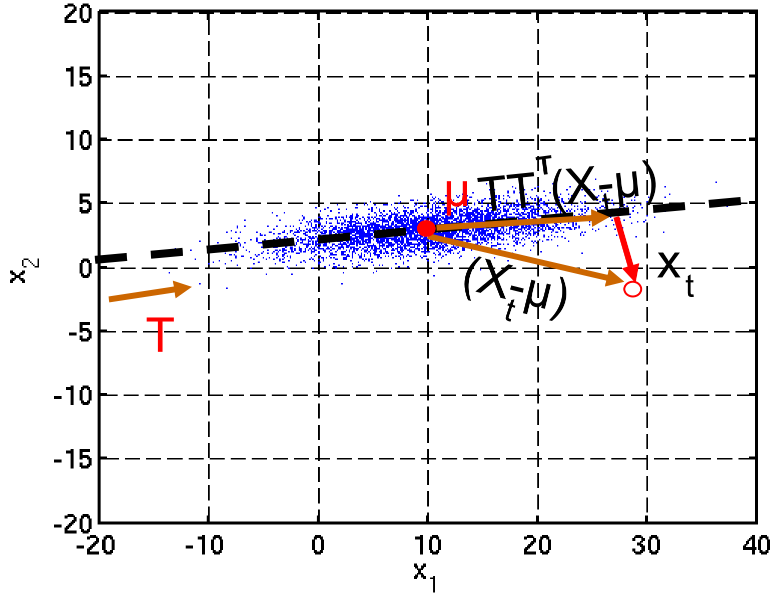
Topological dimension reduction

(*H. 07*): for a given time series x_t look for a minimum of the *reconstruction error*



Topological dimension reduction

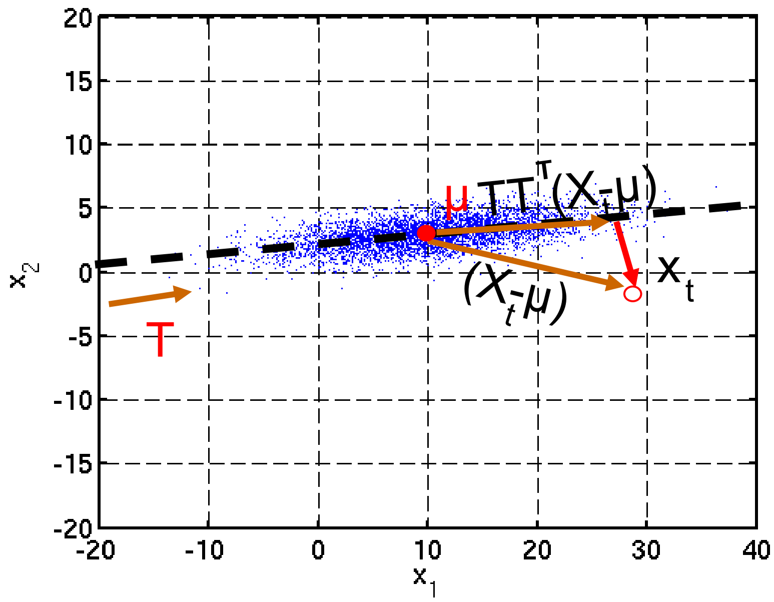
(*H. 07*): for a given time series x_t look for a minimum of the *reconstruction error*



Topological dimension reduction

(*H. 07*): for a given time series x_t look for a minimum of the *reconstruction error*

$$\left\| (x_t - \mu) - \mathbf{T}\,\mathbf{T}^{\mathsf{T}} (x_t - \mu) \right\|_2^2$$



Analysis of embedded data

PCA + Takens' Embedding (Broomhead&King, 1986)

Let $\{z_t\}_{t=1,\ldots,T}, z_t \in \mathbf{R}^d$. Define a new variable $\{x_t^q\} = \{z_t, z_{t-1}, \ldots, z_{t-q}\}$. Let the attractor from Takens' Theorem be in a *linear manifold* defined via $\mathbf{T} \in \mathbf{R}^{nxm}, m << d$. Then

$$\sum_{t=1}^{T} \left\| x_t - \mathbf{T}\mathbf{T}^{\mathsf{T}} x_t \right\|_2^2 \to \min$$

$$\mathbf{T}^{\mathsf{T}}\mathbf{T} = Id,$$

Reconstruction from m *essential coordinates*: $\mathbf{T}\,\mathbf{T}^{\mathsf{T}} x_t$

PCA+Takens: data compression/reconstruction

1.

$$z = \left\{ z_1,\, z_2,\, z_3,\, \ldots \right\}$$

Embedding →

2.

$$x = \begin{Bmatrix} z_1 & z_2 & z_3 & \ldots \\ z_2 & z_3 & z_4 & \ldots \\ z_3 & z_4 & z_5 & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ z_q & z_{q+1} & z_{q+2} & \ldots \end{Bmatrix}$$

↓ Projection

3.

$$x_p(t) = \mathbf{T}^{\mathsf{T}} x(t)$$

← ReconstructionI

4.

$$x_r(t) = \mathbf{T} x_p(t)$$

$$x_r = \begin{Bmatrix} z_1^{(1)} & z_2^{(2)} & z_3^{(3)} & \ldots \\ z_2^{(1)} & z_3^{(2)} & z_4^{(3)} & \ldots \\ z_3^{(1)} & z_4^{(2)} & z_5^{(3)} & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ z_q^{(1)} & z_{q+1}^{(2)} & z_{q+2}^{(3)} & \ldots \end{Bmatrix}$$

ReconstructionII →

5.

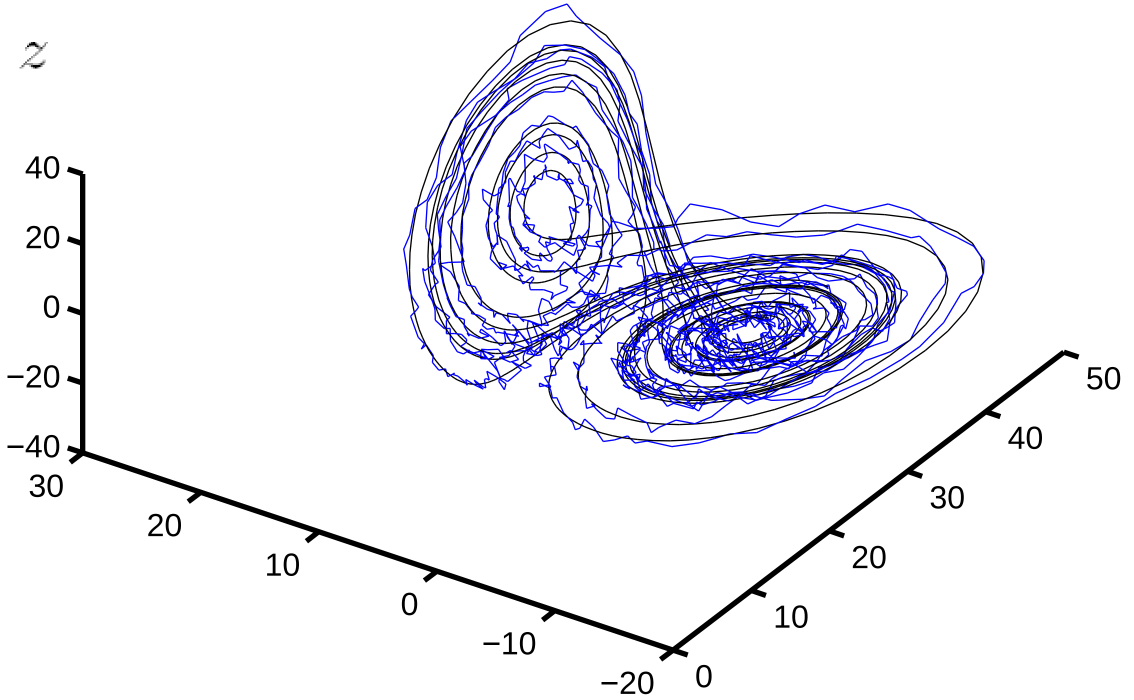
$$z_r = \left\{ z_1^{(1)},\, \tfrac{1}{2}\left(z_2^{(2)} + z_2^{(1)} \right),\, \ldots \right\}$$

$$z_r^{(t)} = \frac{1}{\min\{q,t\}} \sum_{j=0}^{\min\{q-1,t-1\}} z_t^{(t-j)}$$

Example: Lorenz

Lorenz-Oszillator with measurement noise

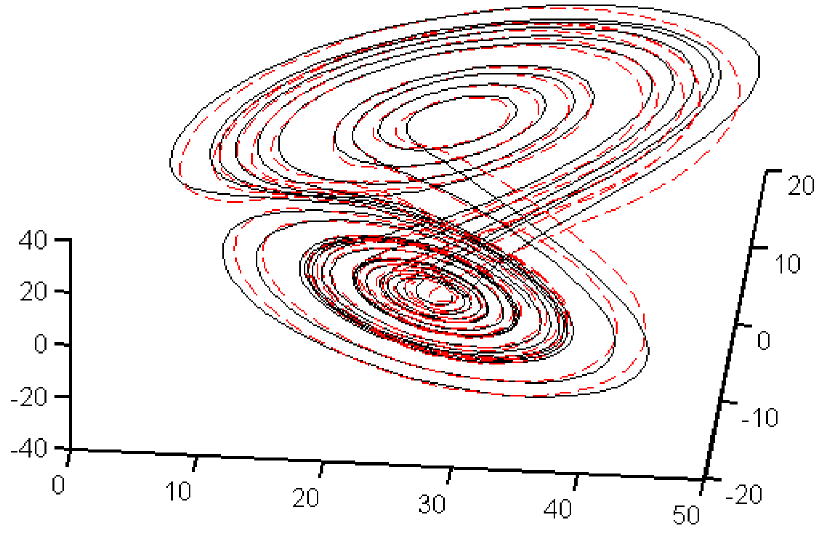
$$
\begin{aligned}
\dot{x} &= \frac{8}{3}x + yz \\
\dot{y} &= -10y + 10z \\
\dot{z} &= -xy + 28y - z
\end{aligned}
$$

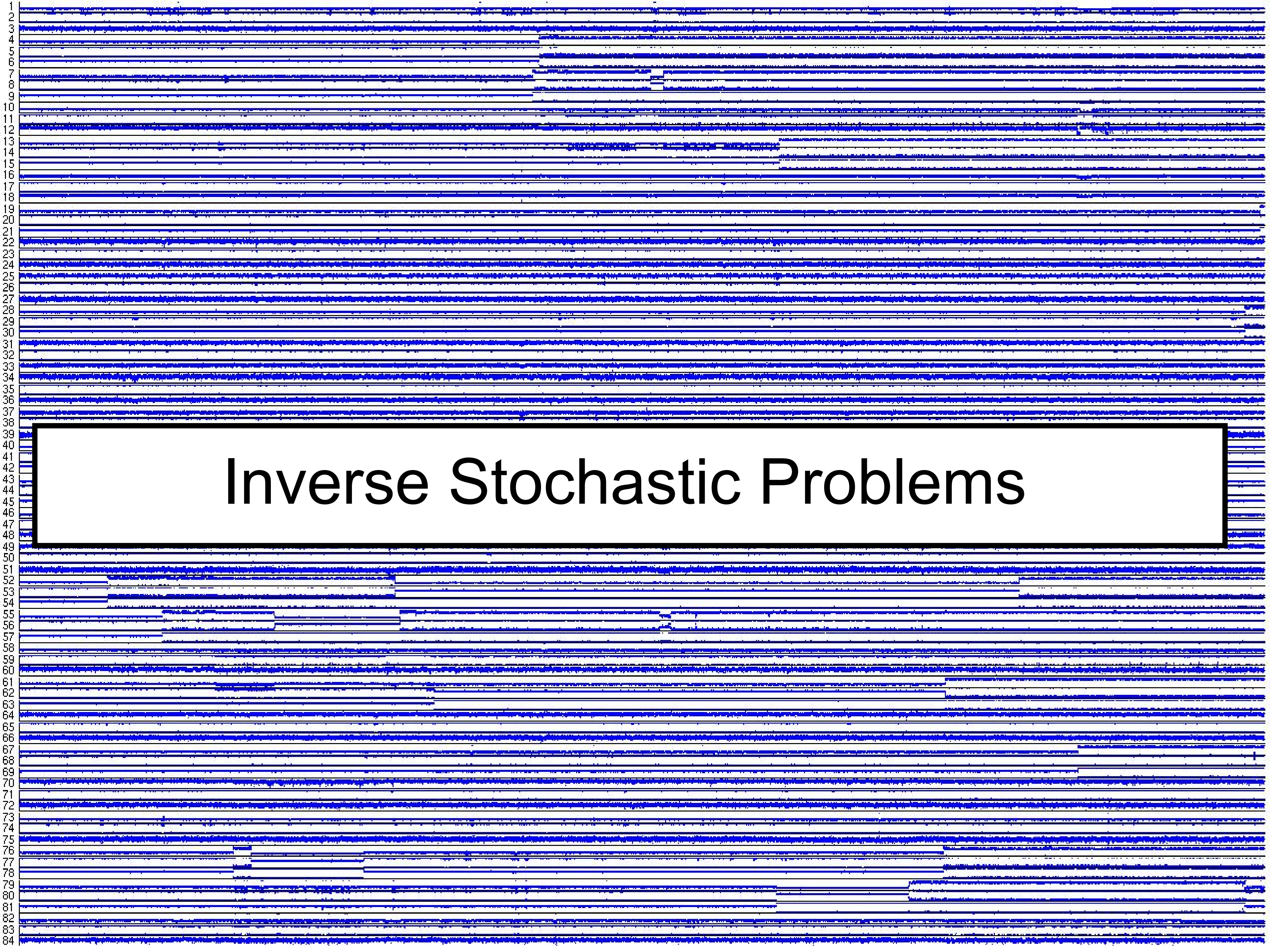


Example: Lorenz

$$\mathbf{T}\ \mathbf{T}^{\mathsf{T}} x_t$$

Reconstructed trajectory
(*red*, for *m=2*)





Inverse Stochastic Problems

And Again: Memo I

A probability space is a measure space such that the measure of the whole space is equal to 1.

In other words: a probability space is a triple $(\Omega, \mathcal{F}, P)$ consisting of a set Ω (called the sample space), a σ-algebra (also called σ-field) $\mathcal{F}$ of subsets of Ω (these subsets are called events), and a measure P on $(\Omega, \mathcal{F})$ such that $P(\Omega) = 1$ (called the probability measure).

Event	Probability
A	$P(A) \in [0,1]$
not A	$P(A') = 1 - P(A)$
A or B	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ $= P(A) + P(B) \quad \text{if A and B are mutually exclusive}$
A and B	$P(A \cap B) = P(A\|B)P(B)$ $= P(A)P(B) \quad \text{if A and B are independent}$
A given B	$P(A\|B)$

Markov Processes: Log-Likelihood

Observed Time Series: $\{X_1, \dots, X_T\}$

Markov-Property:

$$\mathbb{P}\left(X_t | X_1, X_2, \dots, X_{t-1}\right) = \mathbb{P}\left(X_t | X_{t-1}\right)$$

State-Discrete (Markov Chains)

$$\mathbb{P}\left(X_t = s_j | X_{t-1} = s_i\right) = P_{ij}(t-1)$$

State-Discrete (AR(1))

$$X_{t+\tau} = X_t + \tau f(X_t) + \sigma\sqrt{\tau}\epsilon_t$$

$$P(X_{t+\tau}|X_t) = \mathrm{N}(X_t + \tau f(X_t), \tau\sigma^2)$$

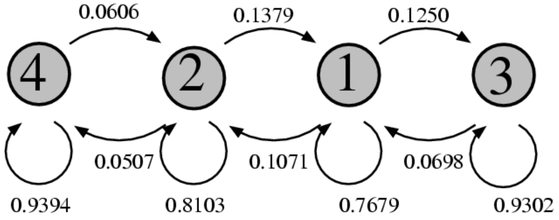
Markov Processes: Log-Likelihood

Observed Time Series: $\{X_1, \ldots, X_T\}$

Markov-Property:

$$\mathbb{P}\left(X_t | X_1, X_2, \ldots, X_{t-1}\right) = \mathbb{P}\left(X_t | X_{t-1}\right)$$

State-Discrete (Markov Chains)



$$\mathbb{P}\left(X_t = s_j | X_{t-1} = s_i\right) = P_{ij}(t-1)$$

State-Discrete (AR(1))

$$X_{t+\tau} = X_t + \tau f(X_t) + \sigma\sqrt{\tau}\epsilon_t$$

$$P(X_{t+\tau}|X_t) = \mathbf{N}(X_t + \tau f(X_t), \tau\sigma^2)$$

Markov Chains: Log-Likelihood

Observed Time Series: $\{X_1, \ldots, X_T\}, \; X_t \in s_1, \ldots, s_m$

Markov-Property:

$$\mathrm{P}\left[X_t = s_j | X_1, X_2, \ldots, X_{t-1} = s_i\right] = \mathrm{P}\left[X_t = s_j | X_{t-1} = s_i\right] = P_{ij}(t)$$

Probability of Observed Time Series (Likelihood):

$$\sum_{j=1}^{m} P_{ij}(t) = 1, \quad \text{for all } t, i$$

$$P_{ij}(t) \geq 0, \quad \text{for all } t, i, j$$

$$\mathrm{P}\left[X_1, \ldots, X_T\right] = \mathrm{P}\left[X_1\right] \prod_{i,j=1}^{m} \prod_{t \in \{t_{ij}\}} P_{ij}(t)$$

Markov Chains: Log-Likelihood

Observed Time Series: $\{X_1, \ldots, X_T\}, \; X_t \in s_1, \ldots, s_m$

Markov-Property:

$$\mathrm{P}\left[X_t = s_j | X_1, X_2, \ldots, X_{t-1} = s_i\right] = \mathrm{P}\left[X_t = s_j | X_{t-1} = s_i\right] = P_{ij}(t)$$

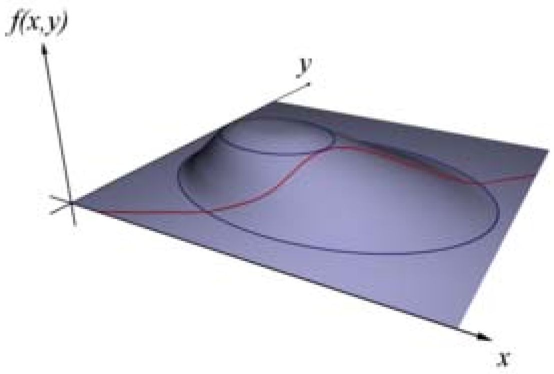
Log-Likelihood Functional:

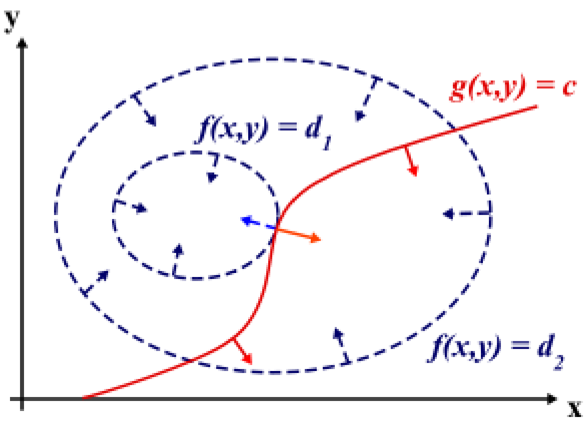
$$\begin{aligned} \sum_{j=1}^{m} P_{ij}(t) &= 1, \quad \text{for all } t, i \\ P_{ij}(t) &\geq 0, \quad \text{for all } t, i, j \end{aligned}$$

$$\begin{aligned} \mathbf{L}\left(P(t)\right) &= \log \mathrm{P}\left[X_1, \ldots, X_T\right] \\ &= \log \mathrm{P}\left[X_1\right] + \sum_{i}^{m} \sum_{j=1}^{m} \sum_{t \in \{t_{ij}\}} \log P_{ij}(t) \rightarrow \max_{P(t)} \end{aligned}$$

Maximization problem is ill-posed

Memo III: Optimization with Constraints





$$f(x,y) = d_n$$
$$g(x,y) = c$$

$$\Rightarrow \quad \nabla f = \lambda \nabla g$$

Lagrange Principle

Take-Home Messages:

1. Numerics *of ODEs and PDEs* is applicable to

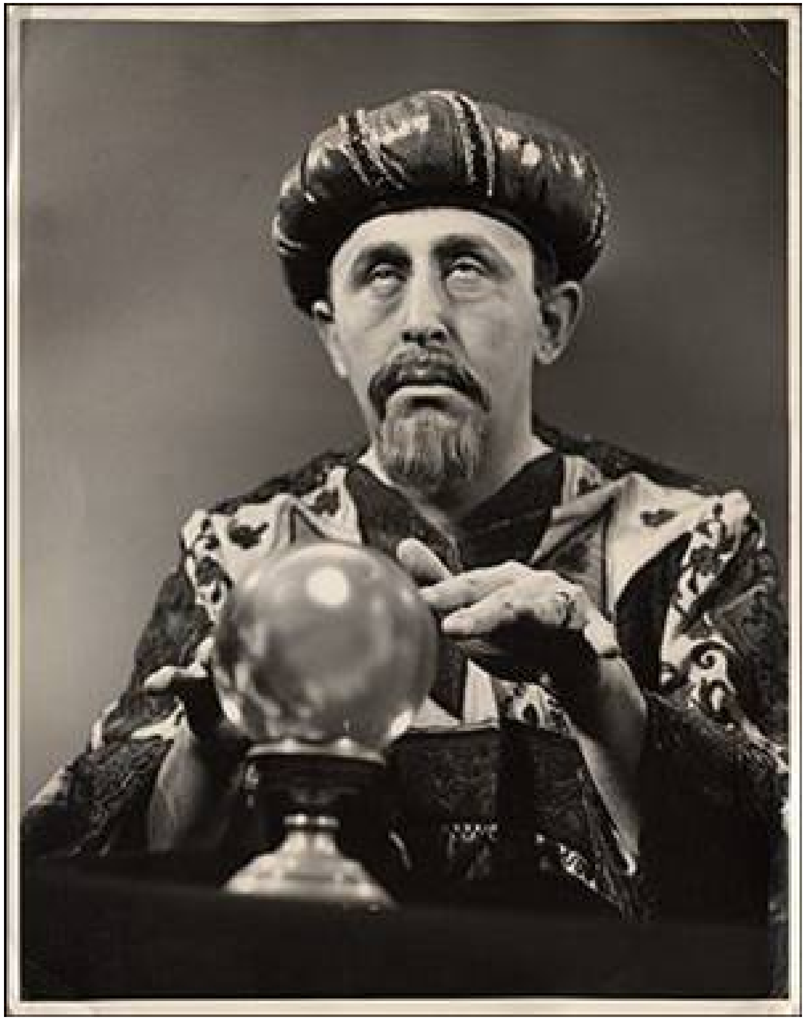
   Stochastic Processes

2. Numerical *inverse problems* in *stochastics* can be understood

   as deterministic minimization problems with constraints

3. *Minimization problems* can be *ill-posed*: additional

   information/assumptions may become necessary



Thank you for attention!