# Audio texture modeling and synthesis using Gaussian process models

MAKSIM KOPTELOV

# What is a definition of texture?

- No universal valid definition
- Texture should exhibit similar characteristics over time
- Can have local structure and randomness, but characteristics remain constant

# Why is it important?

Application in audio processing and sound engineering:

- audio restoration
- missing parts estimation
- etc.

Application in scoring:

- background music
- game music
- etc.

# What we are doing

Input – audio clip t1 seconds

Output – new audio clip t2 seconds, t2 > 0

- Audio texture can be captured from a clip
- It can be used then to synthesize new audio clips with necessary length of time

**Existing audio textures modeling and synthesis methods:**

- *physics based* – not universal
- *sampling based* – mostly manual
- *learning based methods* – more universal, automatic

# Learning based methods

- high dimensionality – curse of dimensionality

dimensionality reduction process must be applied - infers by Gaussian process

- most of audio textures are not linear – learning algorithm cannot be linear

more flexible model must be found – based of Gaussian process

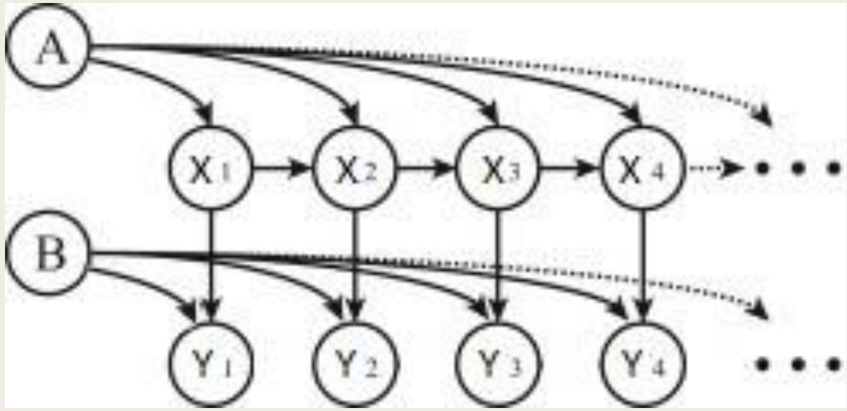
- big computational cost of optimization – necessary to have a good performance

first-order Markov model - based on Gaussian process

Why Gaussian process? Not many approaches can be used for modeling and synthesis

# Gaussian process dynamic model

- latent variable model
- 2 mappings: from a latent space X to the observation space Y and dynamic behavior of latent variables
- kernel function: at least one for mapping from X to Y
- **X – sequence of artificial samples, Y – sequence of real samples**
- A – parameters, B – kernel weights




$Y_i$ is a multivariate Gaussian process indexed by $X_i$ expressed by likelihood

$$p(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{\theta}) = \prod_{t=1}^{N} p(y_t|x_t,\boldsymbol{\theta}) = \frac{1}{(2\pi)^{DN/2}|\boldsymbol{K}_Y|^{D/2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{K}_Y^{-1}\boldsymbol{Y}\boldsymbol{Y}^T\right)\right)$$

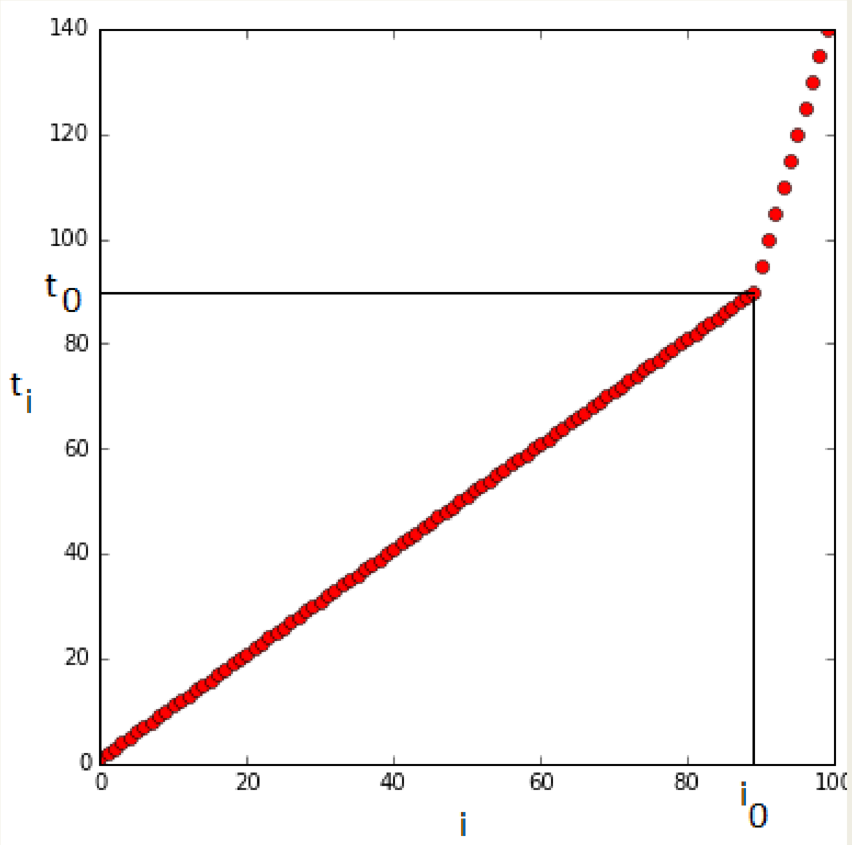
# Gaussian process

**Definition**

Gaussian processes – infinite-dimensional generalization of multivariate normal distributions

Initially we have:

- d random variables ($X_{t1}$,…, $X_{td}$) not equally spaced
- timeline $t_i$
- d = 100
- 100 time indexes $t_i$
- $i_0$ – index when space is increased
- not regular sampling

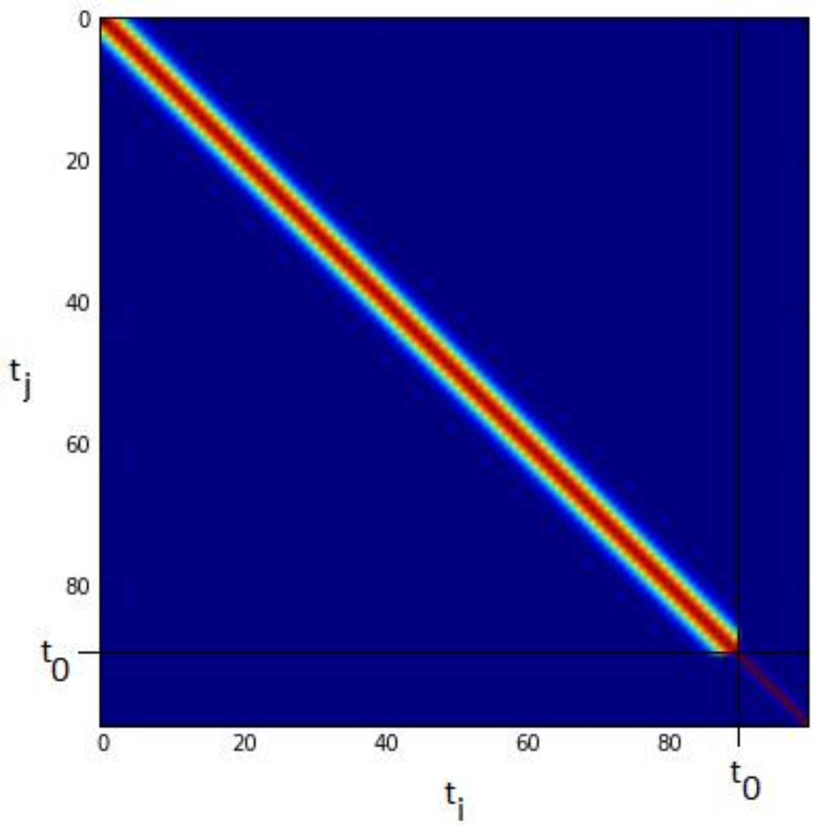


# Gaussian process

**Characterized by mean and covariance**

- all the vector of interest are centered – zero mean initially
- assume a specific form for the covariance of any subsample ($X_{t1}$,…, $X_{td}$)

$$k(t_i, t_j) = cov(X_{t_i}, X_{t_j}) = \alpha \exp\left(-\frac{\|t_i - t_j\|^2}{2l^2}\right)$$




Parameters:

*a* - scale parameter,

*l* - dispersion parameter

# Gaussian process

From one covariance matrix and one sequence of time ($t_1$,...,$t_k$) possible Gaussian distribution ($X_{t1}$,..., $X_{tk}$) with zero mean and covariance C

On the axis:

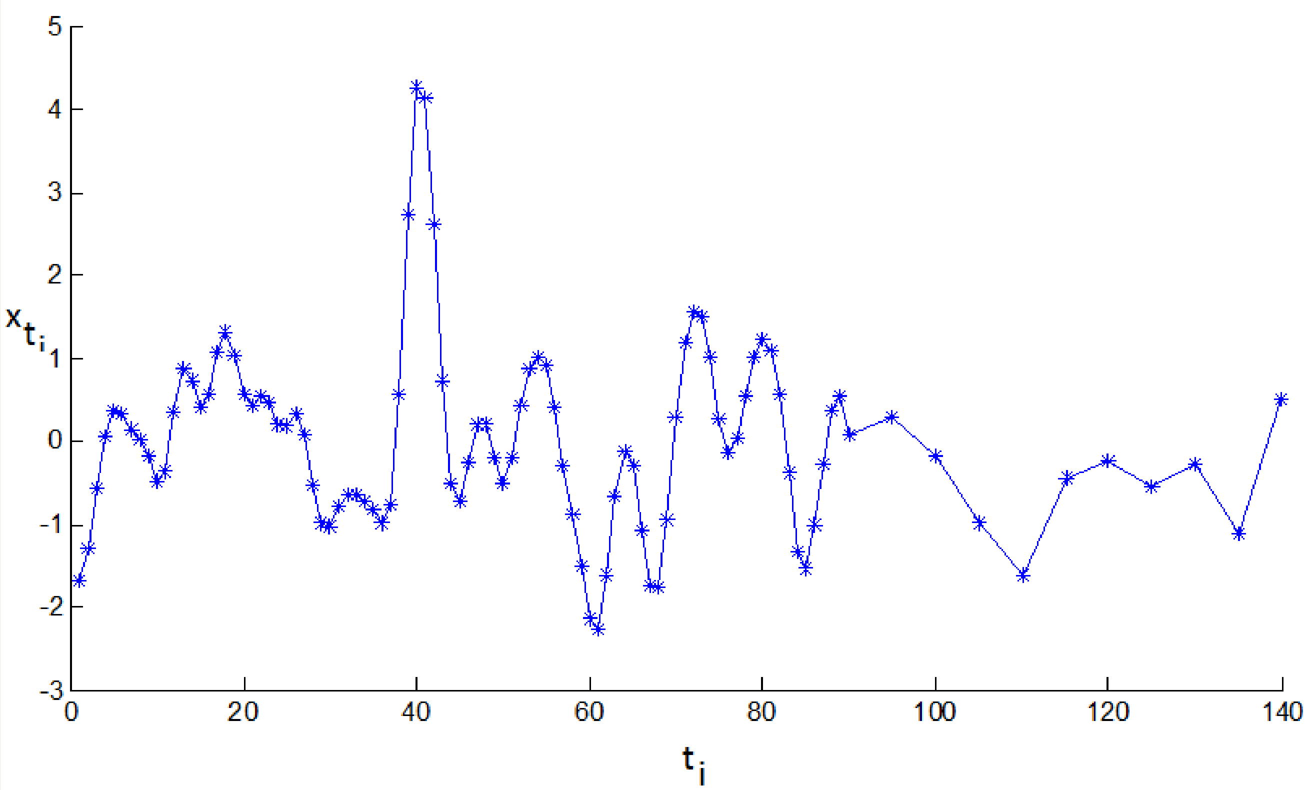
$x_{ti}$ – values of random variables

$t_i$ – time indexes

Trend of samples is not spoiled

by not equally spaced time indexes

and can be easily seen



# References


- N. Lawrence, "Probabilistic Non-linear Principal Component Analysis with Gaussian Process Latent Variable Models", Journal of Machine Learning Research 6 (2005) 1783–1816

- Gerda Strobl, Gerhard Eckel, "SOUND TEXTURE MODELING: A SURVEY", Institute of Electronic Music and Acoustics. University of Music and Dramatic Arts Graz.