Università della Svizzera italiana
Faculty of Informatics
Institute of Computational Science ICS

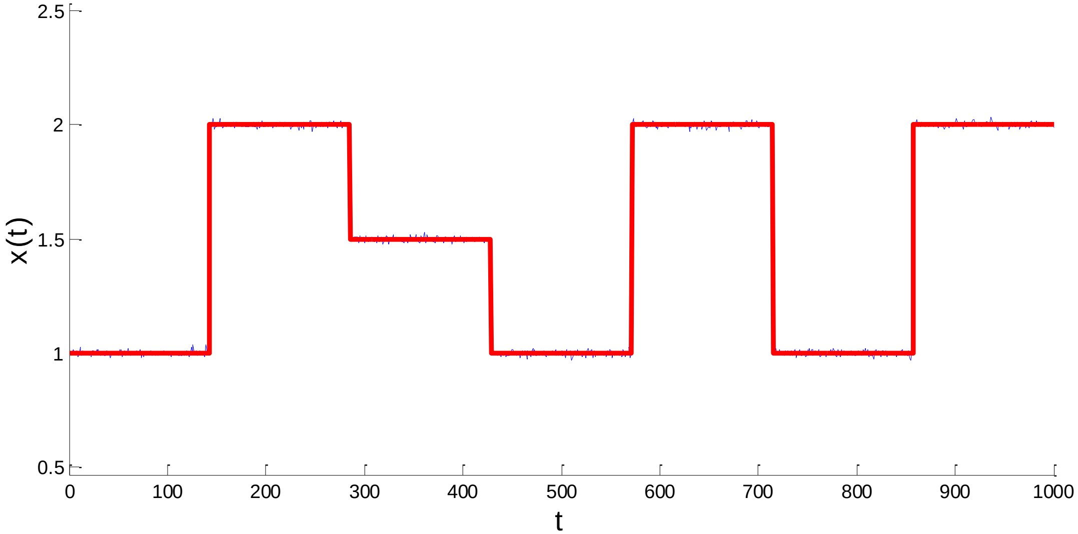
A3: SPA - details

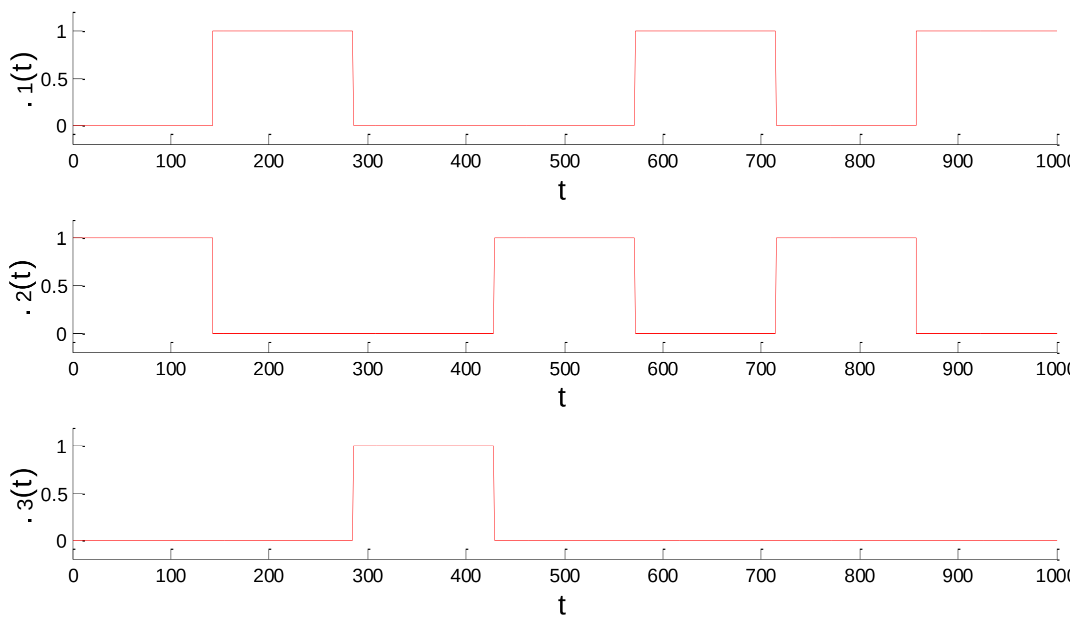
CECAM workshop, Mainz, 2019

Illia Horenko, Lukas Pospisil

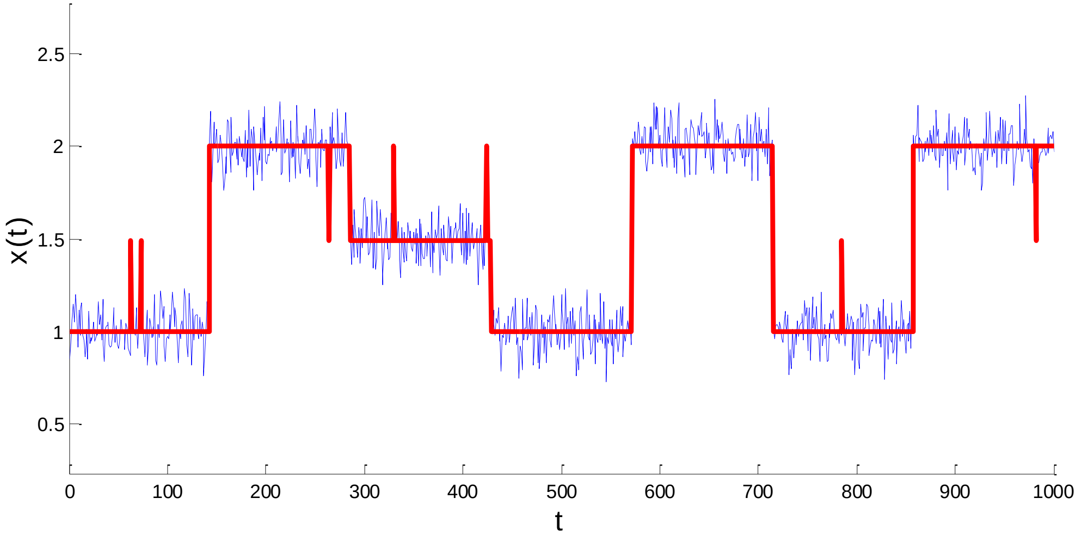
K-means algorithm for time series

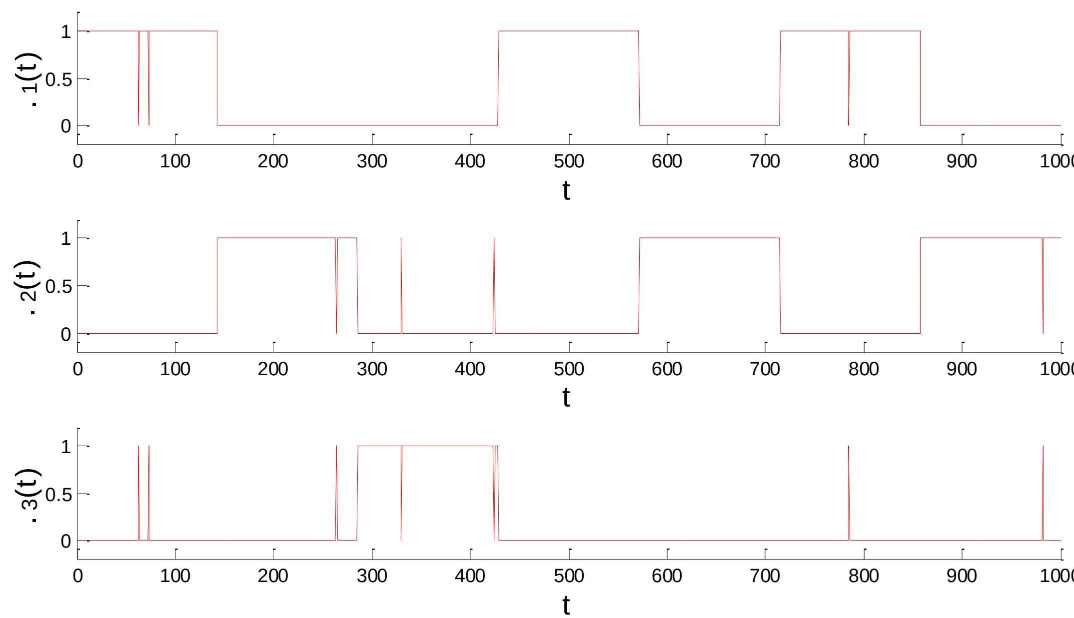
Example - K-means clustering for TS



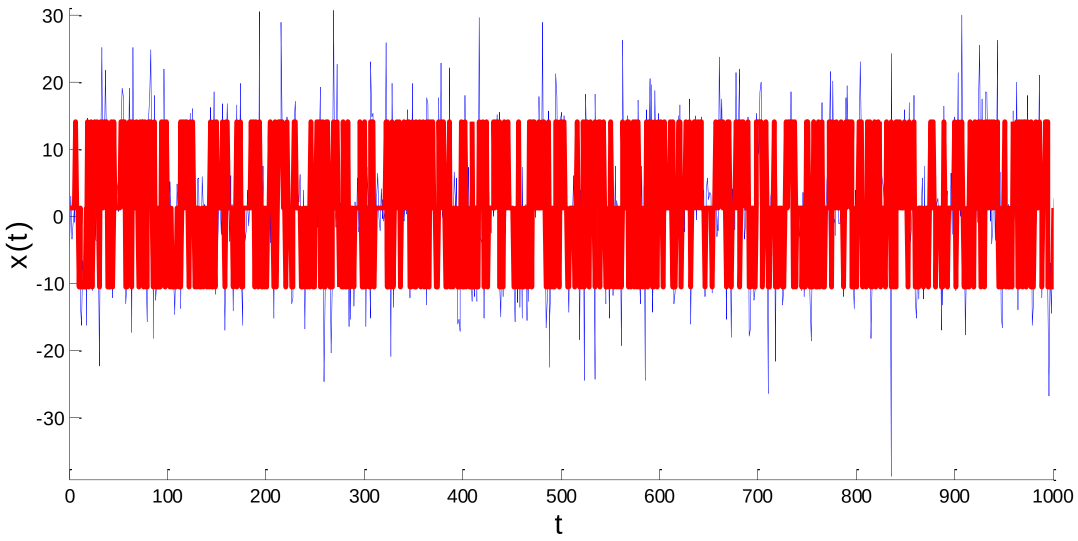


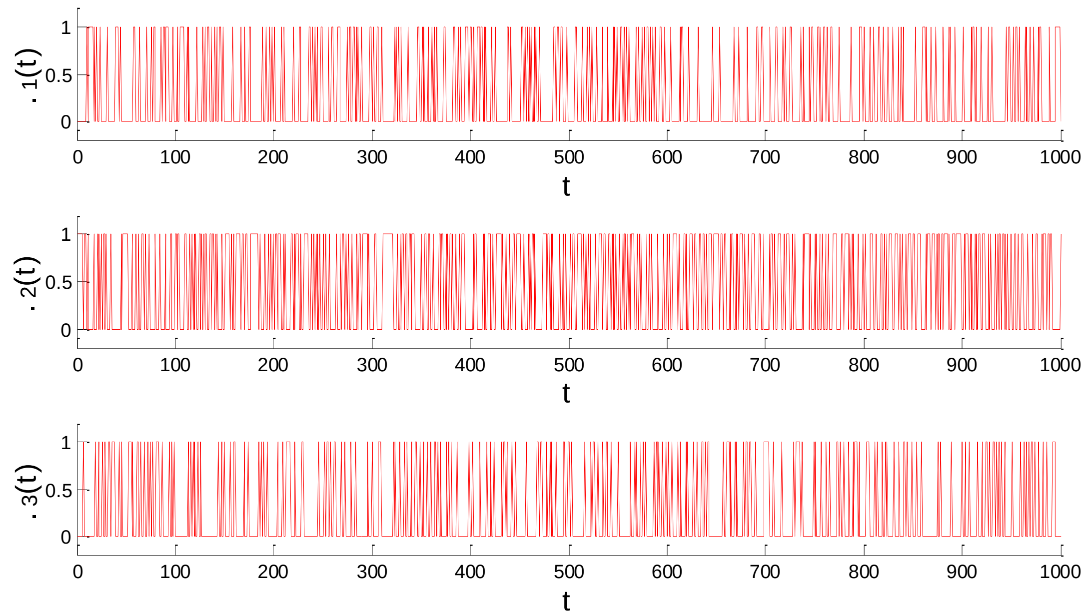
Example - K-means clustering for TS





Example - K-means clustering for TS





Example - K-means clustering for TS

- K-means ignores (doesn't take into account) the time

- we want to implement "rationality" of the cluster changes during time
  (i.e. the time series is not jumping between clusters in crazy way)

$$||\text{change of } \gamma_k(x_t) \text{ during } t|| \quad \leq \quad ||\text{change of } x_t \text{ during } t||$$

caused by the change of cluster (model)

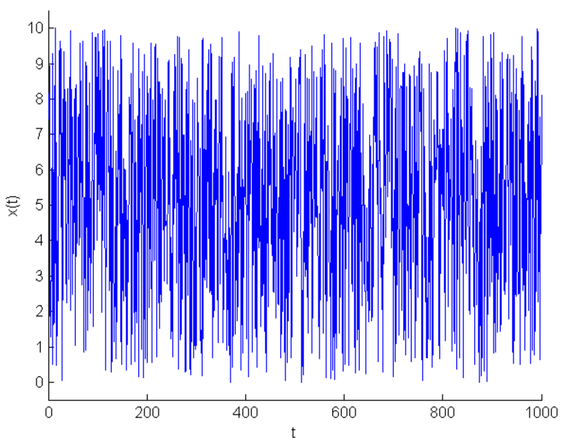
typically caused by the noise

Norm

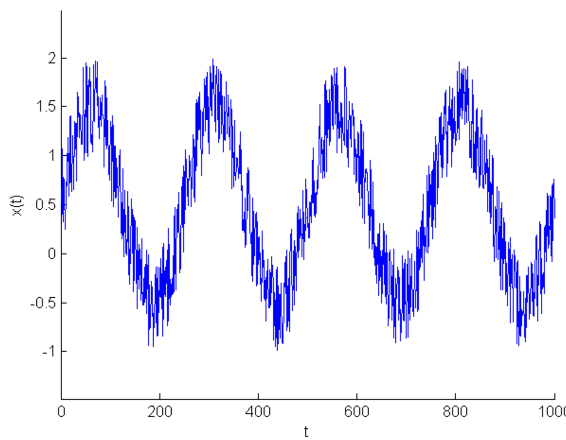
- how to measure the changes of (discrete) function?

the sum of squares of differences between consecutive gamma

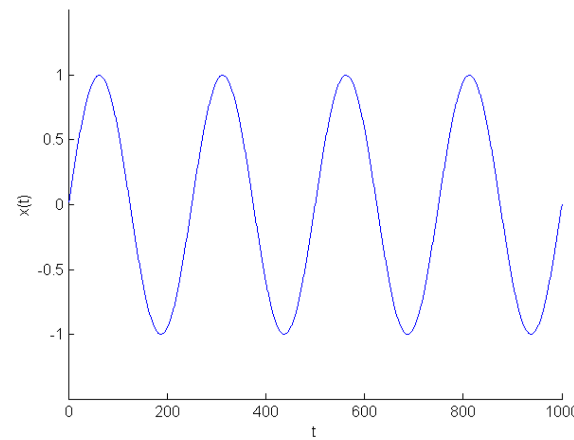
$$\|\gamma_k\|_{\mathcal{H}_1} := \sqrt{\sum_{t=1}^{T-1} (\gamma_k(t+1) - \gamma_k(t))^2} \leq C_k$$



$$\|x\|_{H1} = 1.7453 \cdot 10^4$$



$$\|x\|_{H1} = 153.7537$$

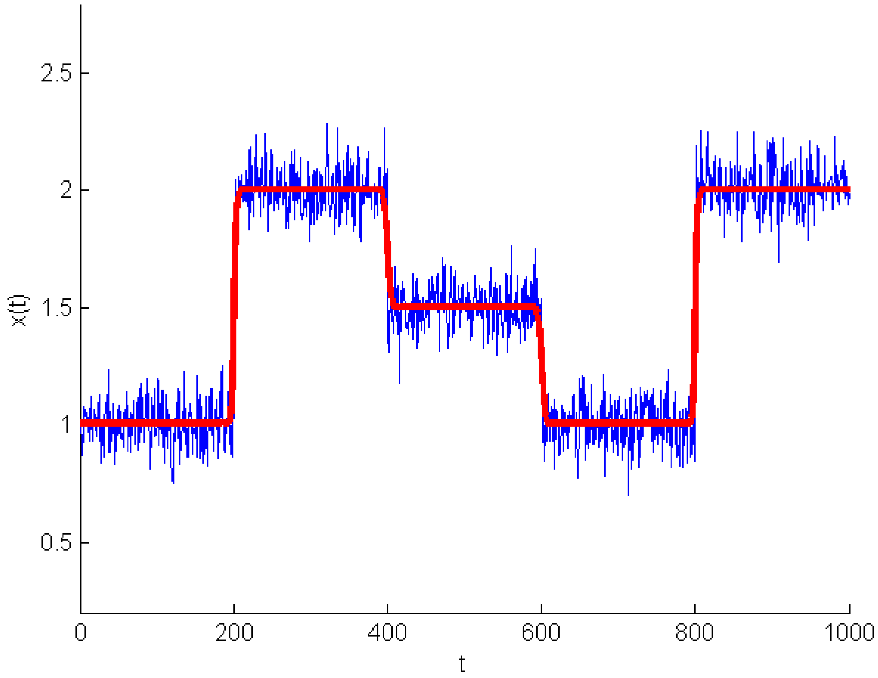


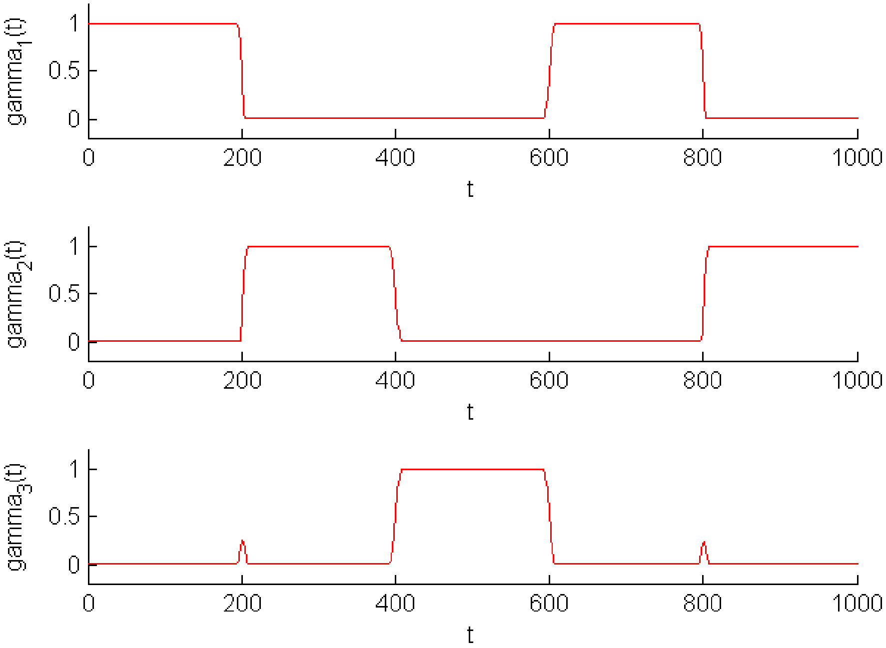
$$\|x\|_{H1} = 0.3152$$

$$[\hat{\theta}, \hat{\Gamma}] = \arg \min_{\theta, \Gamma \in \Omega_\Gamma} \sum_{k=1}^{K} \sum_{t=1}^{T} \gamma_k(t) \cdot \|x(t) - \mu_k\|^2 + \varepsilon^2 \sum_{k=1}^{K} \overbrace{\|\gamma_k\|_{\mathcal{H}_1}^2}^{\gamma_k^T H \gamma_k} \qquad \text{(FEM-H1)}$$

Results

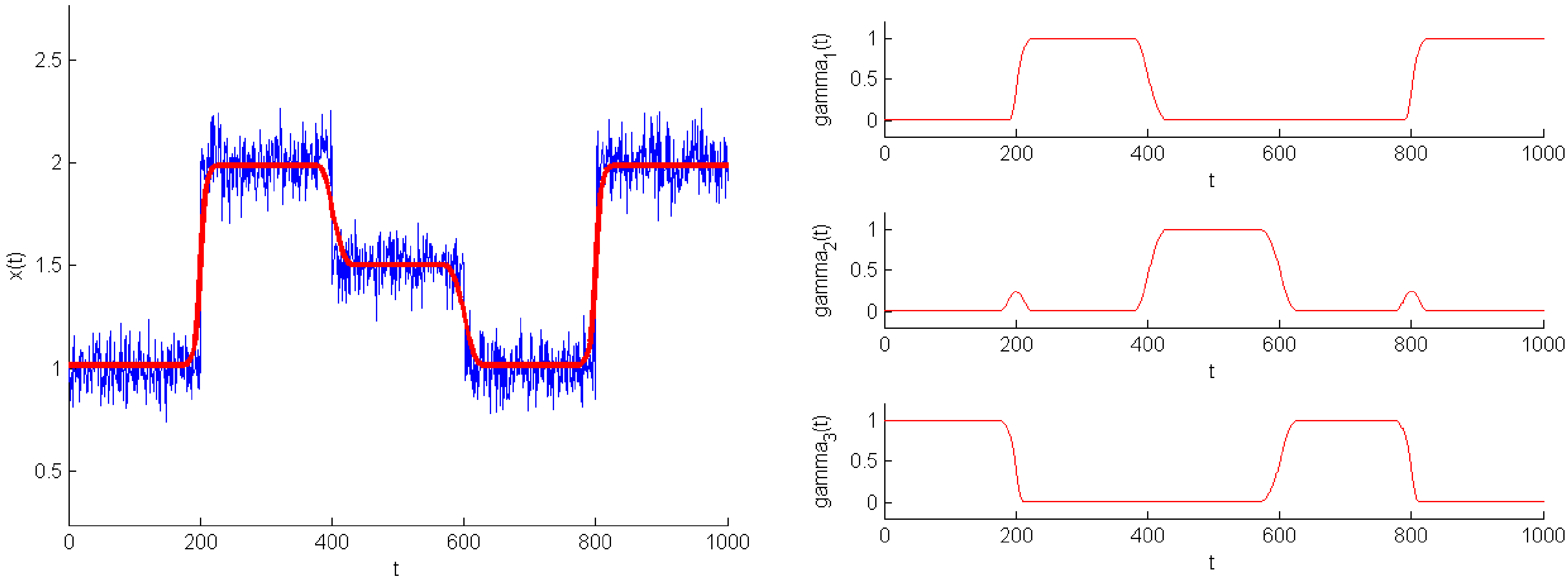
$$\varepsilon^2 = 10$$





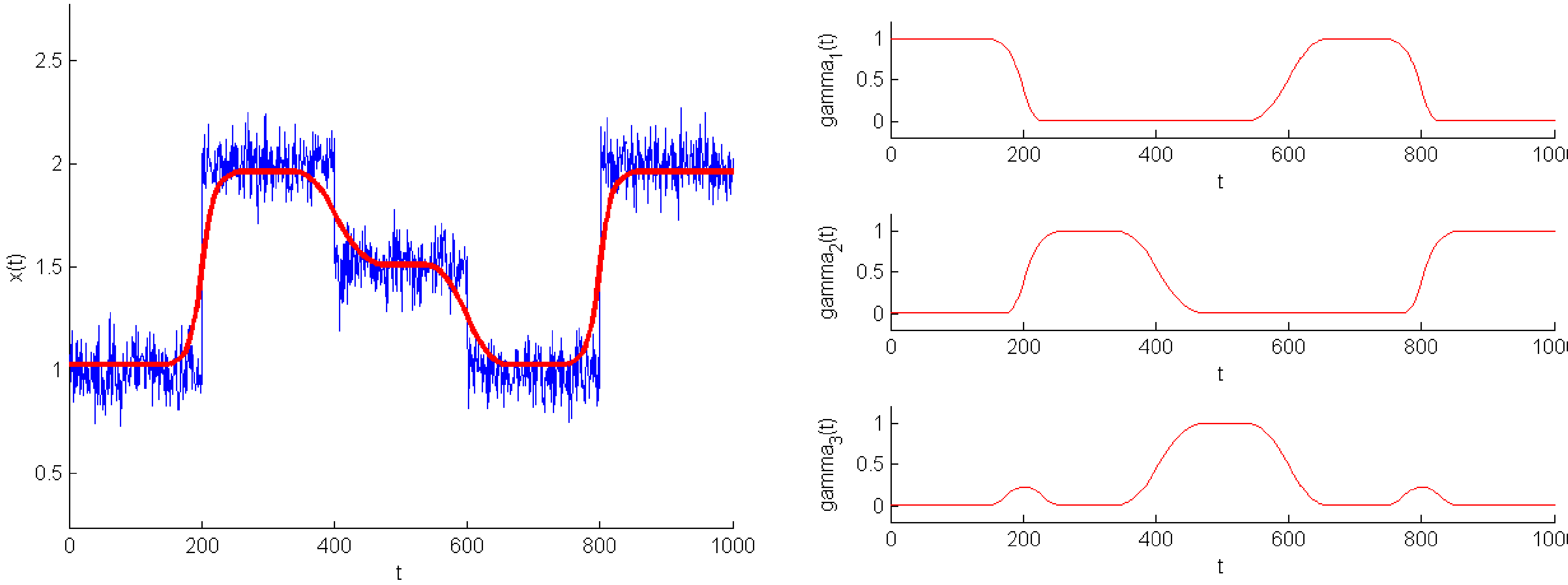
Results

$$\varepsilon^2 = 100$$



Results

$$\varepsilon^2 = 500$$



SPA - details

SPA problem

$$[S^*, \Gamma^*] := \arg\min_{\Gamma \in \Omega_\Gamma} \|X - S\Gamma\|_F^2, \quad \Omega_\Gamma := \left\{\Gamma \in \mathbb{R}^{K,T} \mid \forall t = 1, \ldots, T \; \forall k = 1, \ldots, K : \sum_{k=1}^{K} \Gamma_{k,t} = 1, \Gamma_{k,t} \geq 0\right\}.$$

Set a feasible initial approximation $\Gamma^0 \in \Omega_\Gamma$ *and iteration counter* $\mathrm{it} = 0$.

while $\|L(S, \Gamma^{\mathrm{it}}) - L(S^{\mathrm{it}-1}, \Gamma^{\mathrm{it}-1})\| \geq \varepsilon$

solve $S^{\mathrm{it}} = \arg_S \min L(S, \Gamma^{\mathrm{it}-1})$ *(with fixed* $\Gamma^{\mathrm{it}-1}$*)*

solve $\Gamma^{\mathrm{it}} = \arg_\Gamma \min_{\Gamma \in \Omega_\Gamma} L(S^{\mathrm{it}}, \Gamma)$ *(with fixed* S^{it}*)*

$\mathrm{it} = \mathrm{it} + 1$

endwhile

Return an approximation of the data representation vectors S^{it} *and an approximation of cluster affiliation probability vectors* Γ^{it}.

SPA problem

$$[S^*, \Gamma^*] := \arg\min_{\Gamma \in \Omega_\Gamma} \|X - S\Gamma\|_F^2, \quad \Omega_\Gamma := \ \{\Gamma \in \mathbb{R}^{K,T} \ | \ \forall t = 1, \ldots, T \ \forall k = 1, \ldots, K : \sum_{k=1}^{K} \Gamma_{k,t} = 1, \Gamma_{k,t} \geq 0\}.$$

$$\mathcal{L}(S, \Gamma, \lambda^E, \lambda^I) := \|X - S\Gamma\|_F^2 + \sum_{t=1}^{T} \lambda_t^E \left(\sum_{k=1}^{K} \Gamma_{k,t} - 1 \right) - \sum_{t=1}^{T} \sum_{k=1}^{K} \lambda_{k,t}^I \Gamma_{k,t}$$

SPA problem

$$[S^*, \Gamma^*] := \arg\min_{\Gamma\in\Omega_\Gamma} \|X - S\Gamma\|_F^2, \quad \Omega_\Gamma := \{\Gamma \in \mathbb{R}^{K,T} \mid \forall t = 1,\dots,T\ \forall k = 1,\dots,K : \sum_{k=1}^{K} \Gamma_{k,t} = 1, \Gamma_{k,t} \geq 0\}.$$

$$\mathcal{L}(S,\Gamma,\lambda^E,\lambda^I) := \|X - S\Gamma\|_F^2 + \sum_{t=1}^{T} \lambda_t^E \left(\sum_{k=1}^{K} \Gamma_{k,t} - 1\right) - \sum_{t=1}^{T}\sum_{k=1}^{K} \lambda_{k,t}^I \Gamma_{k,t}$$

$$\begin{aligned}
\nabla_S \mathcal{L}(S,\Gamma,\lambda^E,\lambda^I) = \quad & -2X\Gamma^T + 2S\Gamma\Gamma^T & = 0,\\
\nabla_\Gamma \mathcal{L}(S,\Gamma,\lambda^E,\lambda^I) = \quad & -2S^TX + 2S^TS\Gamma + (\lambda^E)^T \otimes \mathbb{1}_K - \lambda^I & = 0,\\
\nabla_{\lambda_E} \mathcal{L}(S,\Gamma,\lambda^E,\lambda^I) = \quad & \Gamma^T\mathbb{1}_K - \mathbb{1}_T & = 0,\\
\nabla_{\lambda_I} \mathcal{L}(S,\Gamma,\lambda^E,\lambda^I) = \quad & -\Gamma & \leq 0,\\
& \lambda^I & \geq 0,\\
\forall k,t: \quad & \lambda_{k,t}^I \Gamma_{k,t} & = 0,
\end{aligned}$$

SPA problem

$$[S^*, \cancel{\Gamma^*}] := \arg \min_{\cancel{\Gamma \in \Omega_\Gamma}} \|X - S\Gamma\|_F^2, \cancel{\Omega_\Gamma := \{\Gamma \in \mathbb{R}^{K,T} \mid \forall t = 1, \dots, T \; \forall k = 1, \dots, K : \sum_{k=1}^{K} \Gamma_{k,t} = 1, \Gamma_{k,t} \geq 0\}}.$$

$$\mathcal{L}(S, \cancel{\Gamma, \lambda^E, \lambda^I}) := \|X - S\Gamma\|_F^2 \cancel{+ \sum_{t=1}^{T} \lambda_t^E \left(\sum_{k=1}^{K} \Gamma_{k,t} - 1 \right) - \sum_{t=1}^{T} \sum_{k=1}^{K} \lambda_{k,t}^I \Gamma_{k,t}}$$

$$\nabla_S \mathcal{L}(S, \cancel{\Gamma, \lambda^E, \lambda^I}) = -2X\Gamma^T + 2S\Gamma\Gamma^T = 0, \quad \longrightarrow \quad S^* = X\Gamma^T(\Gamma\Gamma^T)^{-1}$$

$$\cancel{\nabla_\Gamma \mathcal{L}(S, \Gamma, \lambda^E, \lambda^I) = -2S^T X + 2S^T S\Gamma + (\lambda^E)^T \otimes \mathbb{1}_K - \lambda^I = 0,}$$

$$\cancel{\nabla_{\lambda_E} \mathcal{L}(S, \Gamma, \lambda^E, \lambda^I) = \Gamma^T \mathbb{1}_K - \mathbb{1}_T = 0,}$$

$$\cancel{\nabla_{\lambda_I} \mathcal{L}(S, \Gamma, \lambda^E, \lambda^I) = -\Gamma \leq 0,}$$

$$\cancel{\lambda^I \geq 0,}$$

$$\cancel{\forall k, t : \lambda_{k,t}^I \Gamma_{k,t} = 0,}$$

SPA problem

$$[S^*, \cancel{\Gamma^*}] := \arg\min_{\cancel{\Gamma\in\Omega_\Gamma}} \|X - S\Gamma\|_F^2, \quad \cancel{\Omega_\Gamma := \{\Gamma \in \mathbb{R}^{K,T} \mid \forall t = 1,\dots,T\ \forall k = 1,\dots,K : \sum_{k=1}^{K} \Gamma_{k,t} = 1, \Gamma_{k,t} \geq 0\}}.$$

$$\mathcal{L}(S, \cancel{\Gamma, \lambda^E, \lambda^I}) := \|X - S\Gamma\|_F^2 \cancel{+ \sum_{t=1}^{T} \lambda_t^E \left(\sum_{k=1}^{K} \Gamma_{k,t} - 1 \right) - \sum_{t=1}^{T} \sum_{k=1}^{K} \lambda_{k,t}^I \Gamma_{k,t}}$$

$$\nabla_S \mathcal{L}(S, \cancel{\Gamma, \lambda^E, \lambda^I}) = -2X\Gamma^T + 2S\Gamma\Gamma^T = 0, \quad \longrightarrow \quad \cancel{S^* = X\Gamma^T(\Gamma\Gamma^T)^{-1}}$$

$$\cancel{\nabla_\Gamma \mathcal{L}(S, \Gamma, \lambda^E, \lambda^I) = -2S^T X + 2S^T S\Gamma + (\lambda^E)^T \otimes \mathbb{1}_K - \lambda^I = 0,}$$

$$\cancel{\nabla_{\lambda_E} \mathcal{L}(S, \Gamma, \lambda^E, \lambda^I) = \Gamma^T \mathbb{1}_K - \mathbb{1}_T = 0,}$$

$$\cancel{\nabla_{\lambda_I} \mathcal{L}(S, \Gamma, \lambda^E, \lambda^I) = -\Gamma \leq 0,}$$

$$\cancel{\lambda^I \geq 0,}$$

$$\cancel{\forall k, t : \lambda_{k,t}^I \Gamma_{k,t} = 0,}$$

$$S^* = X\Gamma^T \left(\Gamma\Gamma^T\right)^+ + \alpha^T R^T, \quad \textit{with parameter } \alpha \in \mathbb{R}^{r,n}, \quad \operatorname{Im} R = \operatorname{Ker} \Gamma^T$$

SPA problem

Using regularization:

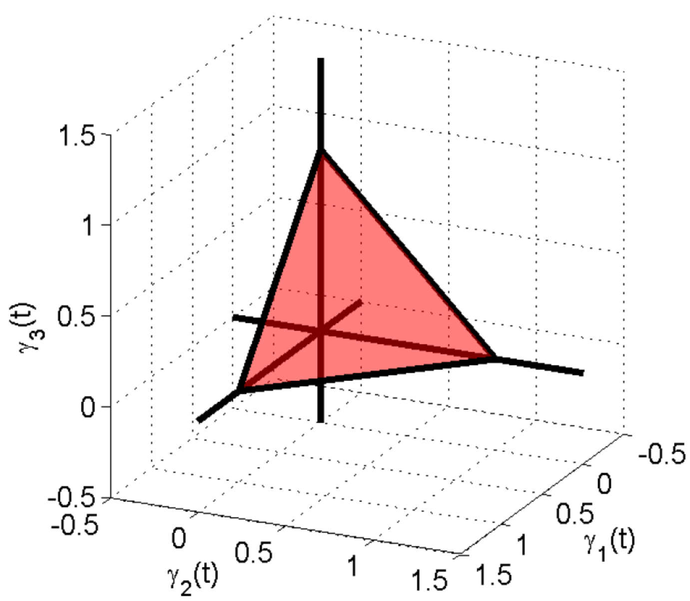
$$\begin{cases} \Phi_S(S) := \dfrac{1}{nK(K-1)} \displaystyle\sum_{i=1}^{n} \sum_{k_1=1}^{K} \sum_{k_2=1}^{K} \left(S_{i,k_1} - S_{i,k_2}\right)^2 \\ S^* = X\Gamma^T H_{\epsilon_S}^{-1}, H_{\varepsilon_S} := \Gamma\Gamma^T + \dfrac{2\varepsilon_S^2}{nK(K-1)}(KI_{K,K} - \mathbb{1}_{K,K}) \end{cases}$$

$$S^* = X\Gamma^T \left(\Gamma\Gamma^T\right)^+ + \alpha^T R^T, \quad \textit{with parameter } \alpha \in \mathbb{R}^{r,n}, \quad \operatorname{Im} R = \operatorname{Ker} \Gamma^T$$

SPA problem

$$\|X - S\Gamma\|_F^2 = \sum_{t=1}^{T} \|x_t - S\gamma_t\|_2^2 = \sum_{t=1}^{T} \left(x_t^T x_t - 2x_t^T S\gamma_t + \gamma_t^T S^T S\gamma_t\right)$$
$$\propto \sum_{t=1}^{T} \tfrac{1}{2}\gamma_t^T (2S^T S)\gamma_t - (S^T x_t)^T \gamma_t.$$

$$0 \leq \gamma_k(t) \leq 1, \quad \forall t : \sum_{k=1}^{K} \gamma_k(t) = 1$$

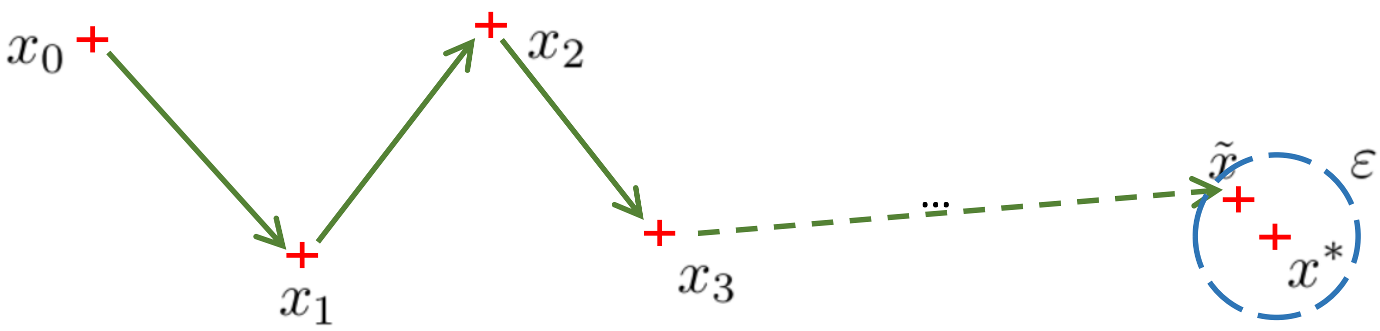


Projected gradient descent methods

(when the projection is known)

Iterative solvers – unconstrained problems

$$x_0 \in \mathbb{R}^n, \quad x_{k+1} = x_k + \alpha_k d_k$$



<u>Gradient descent methods</u> $\quad d_k := -\nabla f(x_k)$

$$x_{k+1} = x_k - \alpha \nabla f(x_k) \qquad \alpha < \frac{2}{\lambda_{\max}}$$

<u>Constant step-size</u>

$$\alpha_k := \arg\min_{\alpha} f(x_k + \alpha d_k) \qquad \alpha_k = \frac{\langle g_k, g_k \rangle}{\langle A g_k, g_k \rangle}$$

<u>Cauchy step-size</u>

Projected gradient descent methods

$$x_0 \in \mathbb{R}^n, \quad x_{k+1} = x_k + \alpha_k d_k$$ (unconstrained)

$$x_0 \in \Omega, \quad x_{k+1} = P(x_k - \alpha_k \nabla f(x_k))$$ (constrained)

$$P(x) = \arg\min_{y \in \Omega} \|x - y\|$$

Projected gradient descent methods

$$x^* = \arg\min_{x\in\Omega} \frac{1}{2}x^T A x - b^T x$$

$$x_0 \in \mathbb{R}^n, \quad x_{k+1} = x_k + \alpha_k d_k$$ (unconstrained)

$$\cancel{x_0 \in \Omega, \quad x_{k+1} = P(x_k - \alpha_k \nabla f(x_k))}$$ (constrained)

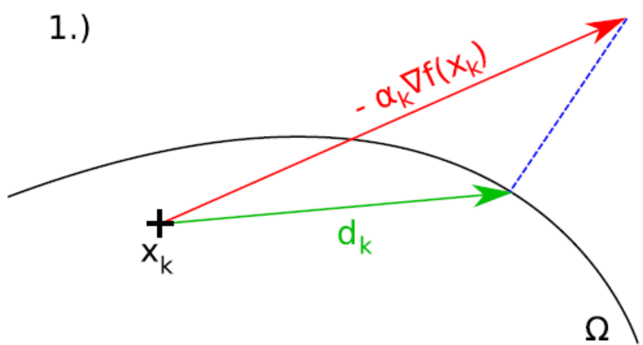
$$P(x) = \arg\min_{y\in\Omega} \|x - y\|$$

$$\boxed{\begin{aligned} & x_0 \in \Omega \\ & d_k = P_\Omega(x_k - \alpha_k \nabla f(x_k)) - x_k \\ & x_{k+1} = x_k + \beta_k d_k \end{aligned}}$$

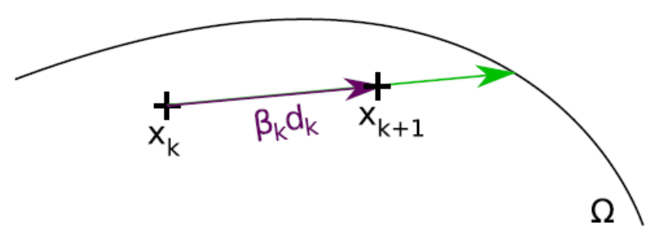
"Projected gradient"

To obtain the decrease of objective function Using "line-search" techniques

1.)

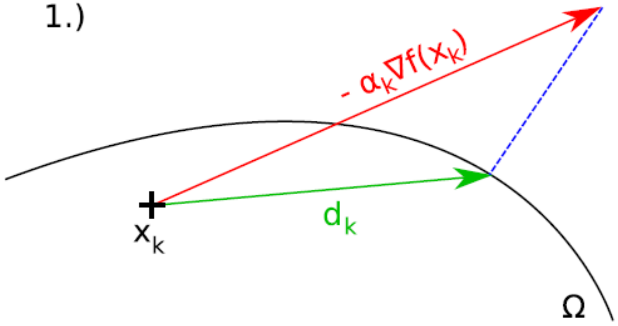


2.)



Example: Spectral Projected Gradient method (SPG)

1.)

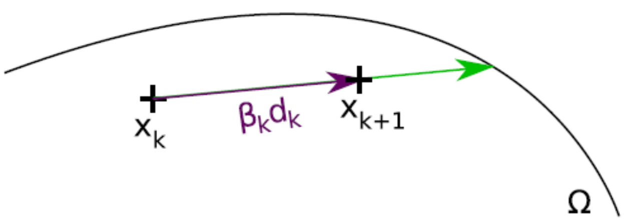


$$d^k = P(x^k - \alpha_k \nabla f(x^k)) - x^k$$

$$P(x) = \arg\min_{y\in\Omega} \|x - y\|$$

α_k - Barzilai-Borwein step-length

2.)



find $\beta_k \in (0,1]$ such that

$$f(x^k + \beta_k d^k) \le f_{\max} + \gamma \beta_k \langle \nabla f(x^k), d^k \rangle$$
$$f_{\max} = \max\{f(x^k), f(x^{k-1}), \dots, f(x^{k-M})\}$$
$$x^{k+1} = x^k + \beta_k d^k$$

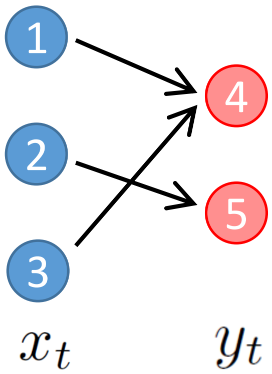
- *Birgin E.G., Raydan M., Martínez J.M.:* Spectral Projected Gradient Methods: Review and Perspectives, *2014*
- *Pospíšil L., Gagliardini P., Sawyer W., Horenko I.*: On a scalable nonparametric denoising of time series signals. *Commun. Appl. Math. Comput. Sci. 13 (2018), no. 1, 107--138. doi:10.2140/camcos.2018.13.107*
- *Pospíšil L.:* Development of Algorithms for Solving Minimizing Problems with Convex Quadratic Function on Special Convex Sets and Applications, *PhD thesis, supervised by Z. Dostál (2015)*

SPA – Markov: example

Benchmark ?

$$x_t \in \{1,2,3\}, \quad t = 1,\dots,T$$
$$y_t \in \{4,5\}$$

$$\Lambda_{exact} := \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

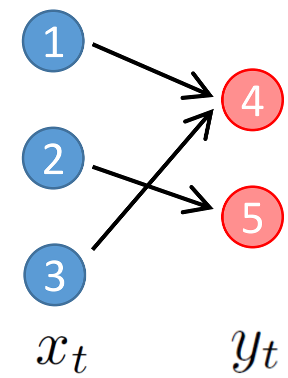


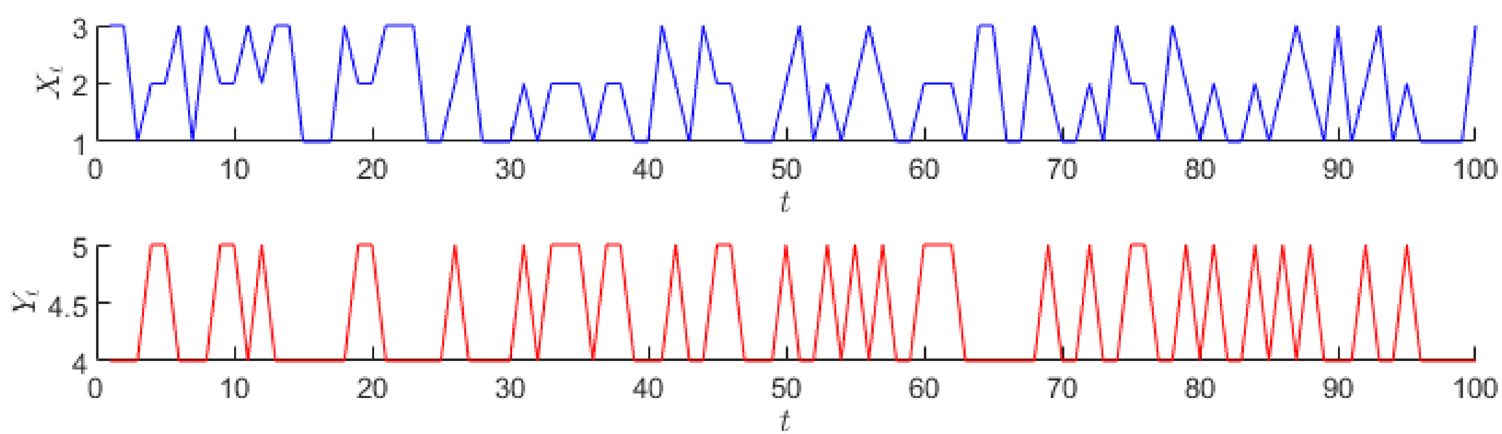
Benchmark ?

$$x_t \in \{1,2,3\}, \quad t = 1,\ldots,T$$
$$y_t \in \{4,5\}$$

$$\Lambda_{exact} := \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

- generate random x_t
- compute Γ^X
- compute $\Gamma^Y_{:,t} = \Lambda_{exact}\Gamma^X_{:,t}$
- set (expected value) $y_t = \Gamma^Y_{1,t} \cdot 4 + \Gamma^Y_{1,t} \cdot 5$



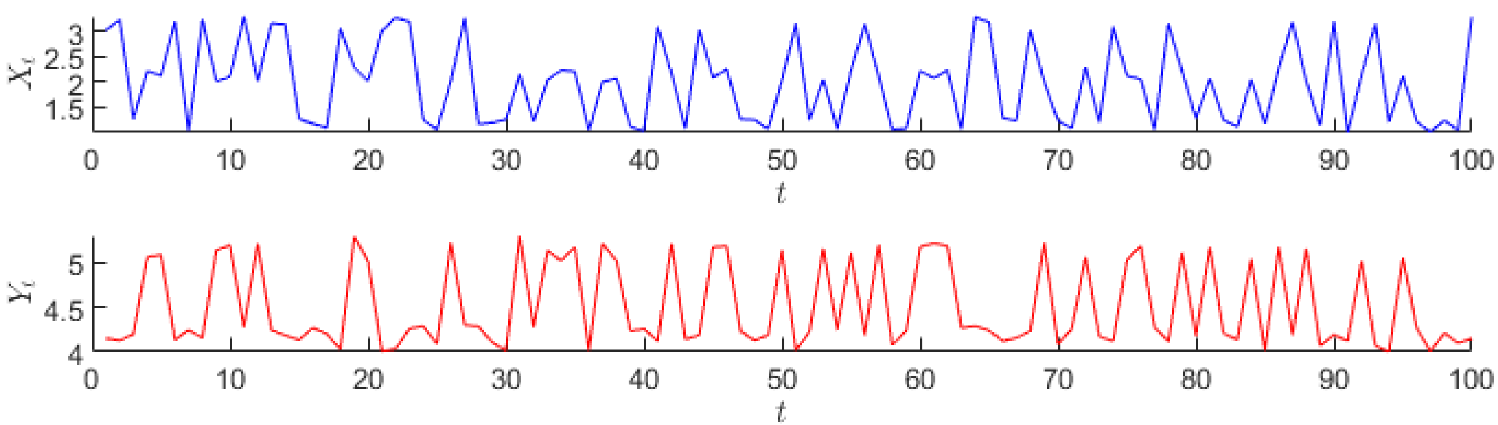


Benchmark ?

- “add” noise (-> continuous data)

$$x_t := x_t + \varepsilon_t^X, \quad \varepsilon_t^X \sim \mathcal{N}(0, \sigma)$$
$$y_t := y_t + \varepsilon_t^Y, \quad \varepsilon_t^Y \sim \mathcal{N}(0, \sigma)$$

- throw away (forget) everything what we know except the data



Benchmark ?

- we have continuous processes x_t and y_t

- what is the relationship between these two processes, i.e.,
  if someone provide “new” x , what will be the most probable y ?

Benchmark ?

- we have continuous processes x_t and y_t

- what is the relationship between these two processes, i.e.,
  if someone provide "new" x, what will be the most probable y ?

- we are looking for conditional probabilities

$$\Gamma^Y_{:,t} = \Lambda \Gamma^X_{:,t}$$

- but we need categorical data (= discretization)

- *"brilliant idea":* what about K-means with

$$\begin{aligned} S^X &= [1,2,3] \\ S^Y &= [4,5] \end{aligned}$$

(Classic approach: discretize data and then estimate conditional probabilities.)

Benchmark ?

Theorem 2. *Let* $x_t \in \mathbb{R}^n$ *and* $y_t \in \mathbb{R}^m$ *be two time series of length* T, $X = [x_1, \ldots, x_T] \in \mathbb{R}^{n,T}, Y = [y_1, \ldots, y_T] \in \mathbb{R}^{m,T}$. *The solution of* (SPA$_2$) *in the form*

$$[S_\varepsilon^*, \Gamma_x^*] = \arg\min_{S,\Gamma} \min_{\Gamma_x \in \Omega_\Gamma} \|X_\varepsilon - S_\varepsilon \Gamma_x\|_F^2 \tag{50}$$

with

$$X_\varepsilon := \begin{bmatrix} Y \\ \varepsilon X \end{bmatrix}, \quad S_\varepsilon := \begin{bmatrix} S_y \Lambda \\ \varepsilon S_x \end{bmatrix}, \tag{51}$$

and $\varepsilon \geq 0$ *is equivalent to the solution of* (SPA$_2$) *problems*

$$[S_x^*, \Gamma_x^*] := \arg\min_{S_x,\Gamma_x} \min_{\Gamma_x \in \Omega_\Gamma} \|X - S_x \Gamma_x\|_F^2, \tag{52}$$

$$[S_y^*, \Gamma_y^*] := \arg\min_{S_y,\Gamma_y} \min_{\Gamma_y \in \Omega_\Gamma} \|Y - S_y \Gamma_y\|_F^2, \tag{53}$$

in Tikhonov-sense with regularization parameter ε *and* $\Lambda \in \mathbb{R}^{K,T}$ *is left-stochastic matrix of conditional probabilities such that the discrete Bayesian and Markovian model equations*

$$\Gamma_y = \Lambda \Gamma_x, \tag{54}$$

are satisfied.

Benchmark ?

Theorem 2. *Let $x_t \in \mathbb{R}^n$ and $y_t \in \mathbb{R}^m$ be two time series of length T, $X = [x_1, \dots, x_T] \in \mathbb{R}^{n,T}$, $Y = [y_1, \dots, y_T] \in \mathbb{R}^{m,T}$. The solution of* (SPA$_2$) *in the form*

$$[S_\varepsilon^*, \Gamma_x^*] = \arg \min_{S,\Gamma} \min_{\Gamma_x \in \Omega_\Gamma} \|X_\varepsilon - S_\varepsilon \Gamma_x\|_F^2 \tag{50}$$

with

$$X_\varepsilon := \begin{bmatrix} Y \\ \varepsilon X \end{bmatrix}, \quad S_\varepsilon := \begin{bmatrix} S_y \Lambda \\ \varepsilon S_x \end{bmatrix}, \tag{51}$$

and $\varepsilon \geq 0$ is equivalent to the solution of (SPA$_2$) *problems*

$$[S_x^*, \Gamma_x^*] := \arg \min_{S_x,\Gamma_x} \min_{\Gamma_x \in \Omega_\Gamma} \|X - S_x \Gamma_x\|_F^2, \tag{52}$$

$$[S_y^*, \Gamma_y^*] := \arg \min_{S_y,\Gamma_y} \min_{\Gamma_y \in \Omega_\Gamma} \|Y - S_y \Gamma_y\|_F^2, \tag{53}$$

in Tikhonov-sense with regularization parameter ε and $\Lambda \in \mathbb{R}^{K,T}$ is left-stochastic matrix of conditional probabilities such that the discrete Bayesian and Markovian model equations

$$\Gamma_y = \Lambda \Gamma_x, \tag{54}$$

are satisfied.

$$\boxed{X, Y \longrightarrow S_x, S_y \Lambda} \qquad \Lambda = ?$$

Benchmark ?

- if someone provide "new" x , what will be the most probable y ?

$$\boxed{X, Y \longrightarrow S_x, S_y \Lambda} \quad \Lambda = ?$$

Benchmark ?

- if someone provide “new” x , what will be the most probable y ?

$$\gamma^X = \arg\min_{\gamma\in\Omega} \|x - \underbrace{S^X}_{}\gamma\|_2$$

$$\boxed{X, Y \longrightarrow S_x\,, S_y \Lambda} \qquad \Lambda = ?$$

Benchmark ?

- if someone provide “new” x , what will be the most probable y ?

$$\gamma^X = \arg\min_{\gamma \in \Omega} \|x - \underbrace{S^X}\gamma\|_2$$

$$y \approx S^Y \gamma^Y$$

$$\boxed{X, Y \longrightarrow S_x, S_y \Lambda}$$

$$\Lambda = ?$$

Benchmark ?

- if someone provide “new” x , what will be the most probable y ?

$$\gamma^X = \arg\min_{\gamma\in\Omega} \|x - \underbrace{S^X}\gamma\|_2$$

$$y \approx S^Y\gamma^Y = \underbrace{S^Y\Lambda}\gamma^X$$

$$\gamma^Y = \Lambda\gamma^X$$

$$\boxed{X, Y \longrightarrow S_x\,, S_y\Lambda} \qquad \Lambda = ?$$