Detecting Changes
in Markov Process

Ce projet est cofinancé par l'Union européenne.
L'Europe s'engage en Normandie
avec le Fonds européen de Développement Régional (FEDER)

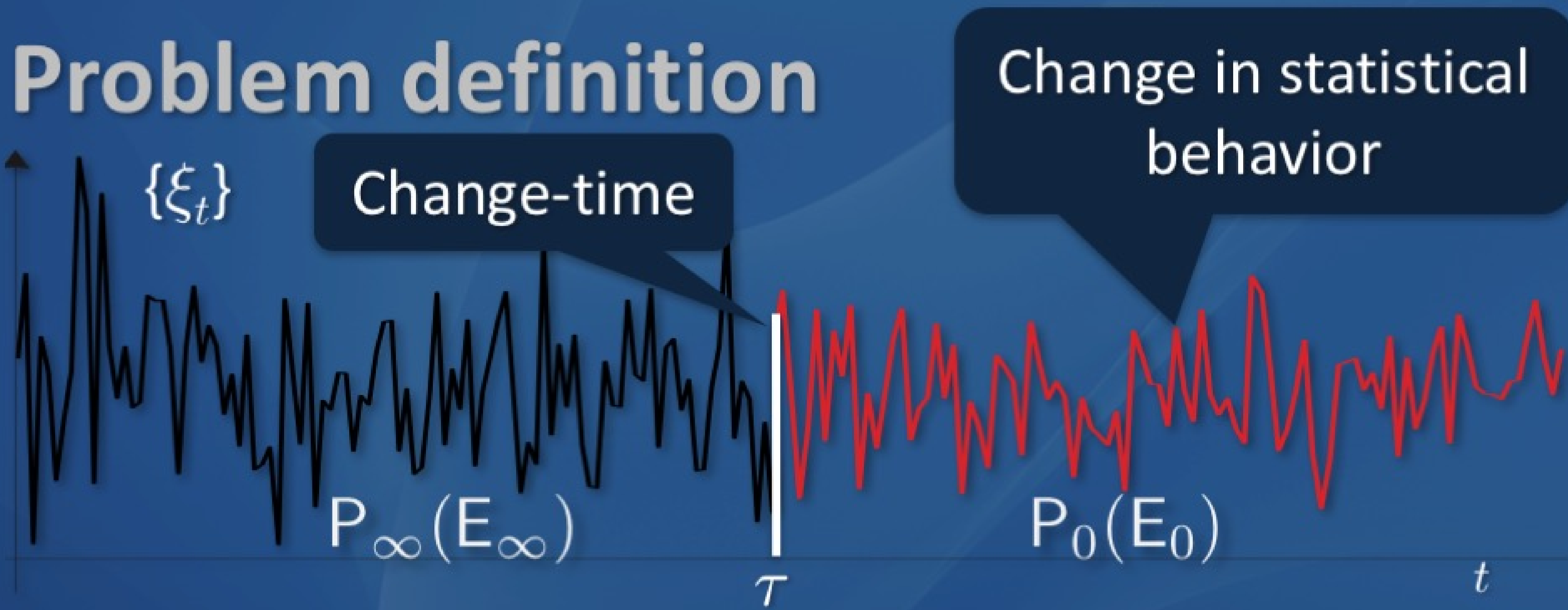
RÉGION NORMANDIE

George V. Moustakides
University of Patras, Greece

Outline

- Problem definition
- Model for change-imposing mechanism
- Performance measures
  - Known change-imposing mechanism
  - Unknown change-imposing mechanism
- Examples (emphasis on Markov processes)

Problem definition



Detect change as soon as possible

Data become available sequentially: at each instant t obtain new sample ξ_t.

Detector: At every time instant t consult available data ξ_1 ,..., ξ_t and use them to decide whether a change took place until and including t.

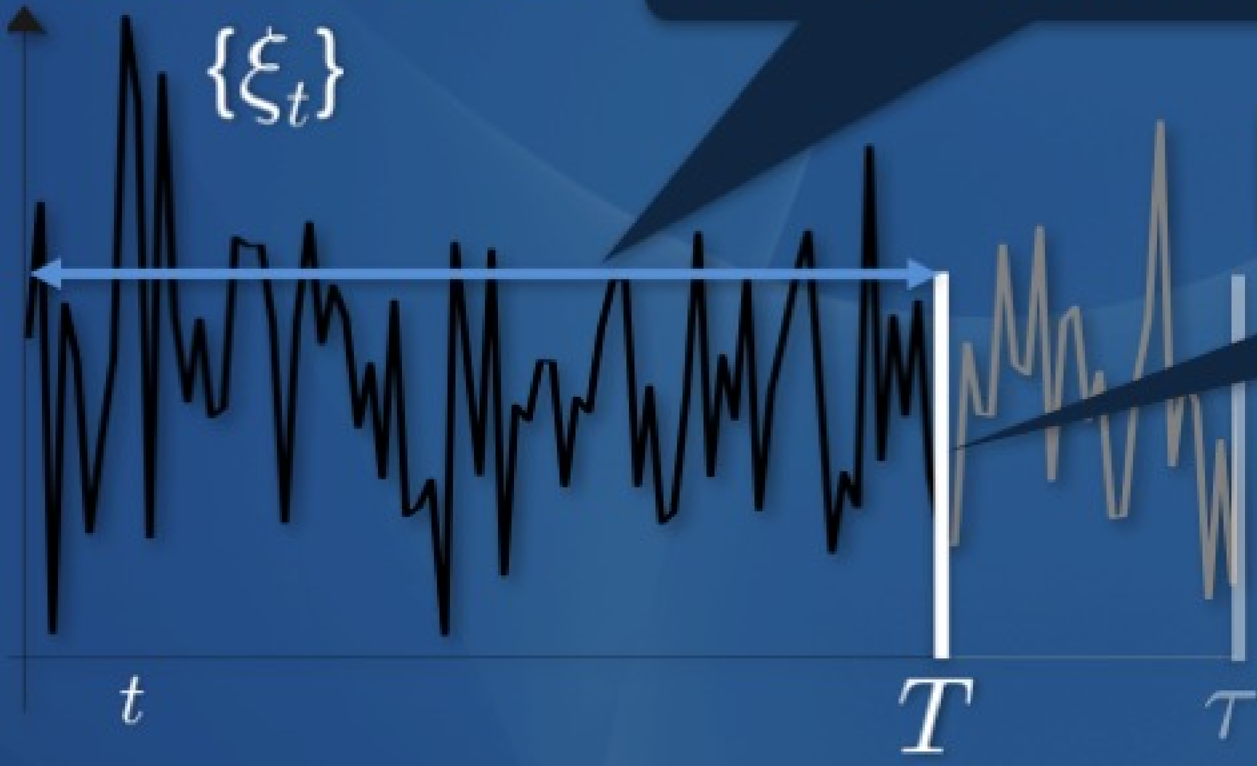
Sequential detector

Each instant t decide between:

- ~~A change took place before and including t.~~ **STOP**
- ~~A change didn't take place before and including t.~~ **Ask for more observations**

Consequently, a sequential detector is simply a **stopping time** T which is adapted to the **observation history** (filtration generated by the observations).

False Alarm



Design T to **optimize** Detection performance while **controlling** False Alarms

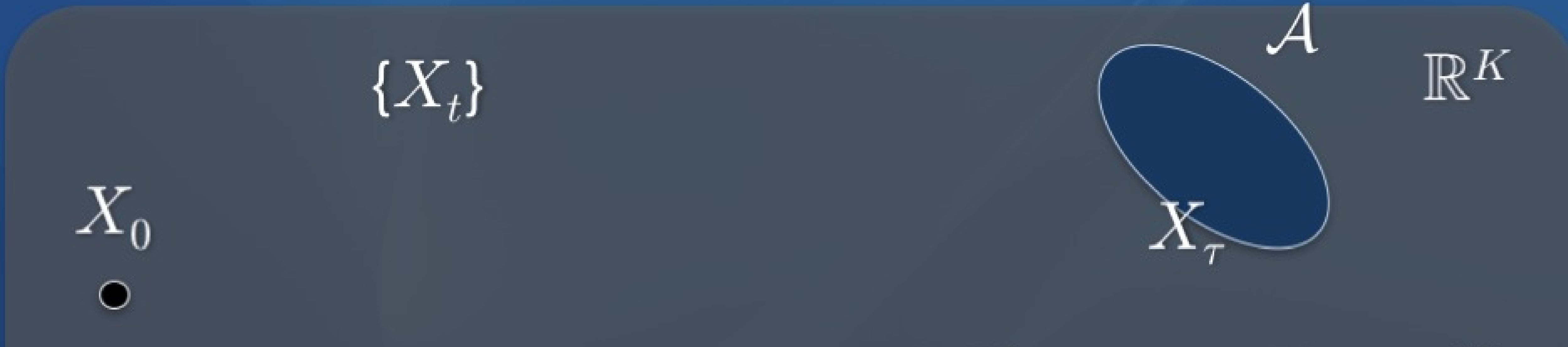
(Successful) Detection

Structural Change-detection in Exchange Rates
Portfolio Monitoring
Electronic Communications
Seismology
Speech & Image Processing (segmentation)
Vibration monitoring (Structural health monitoring)
Security monitoring (fraud detection)
Spectrum monitoring
Scene monitoring
Network monitoring (router failures, attack detection)
⋮

CUSUM: 3,000 hits in 2015. Google Scholar.
80% in Change Detection: 2300 articles

Model for change imposing mechanism

A random vector process $\{X_t\}$ evolves in time in $\mathbb{R}^K$
$\mathcal{A}$ is a subset in $\mathbb{R}^K$



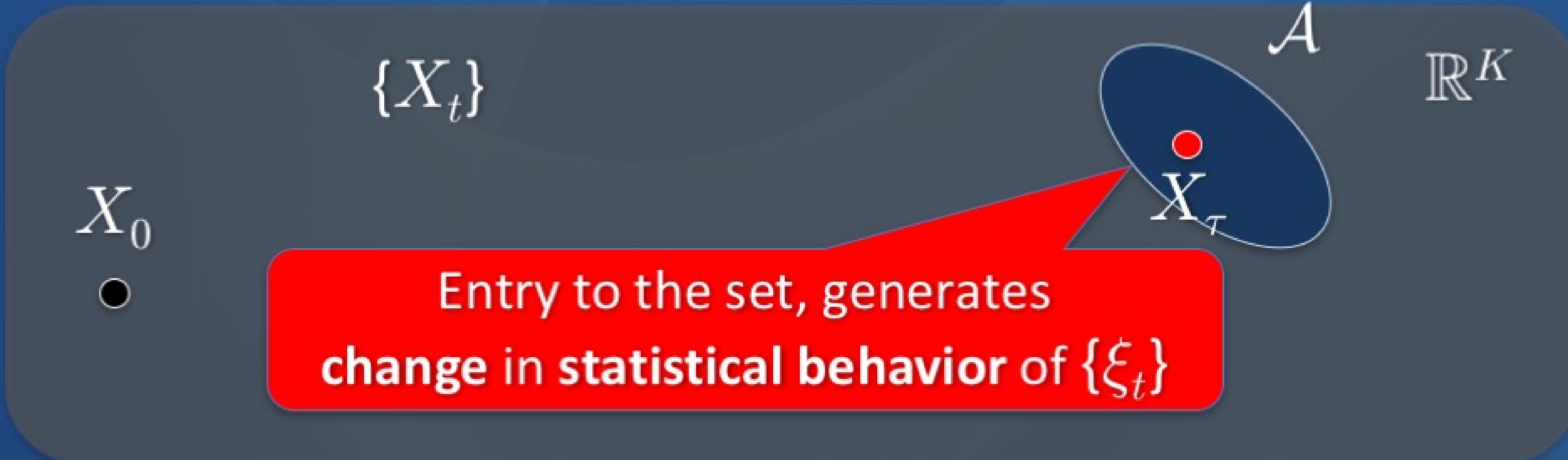
$$\tau = \inf\{t > 0 : \ X_t \in \mathcal{A}\}$$

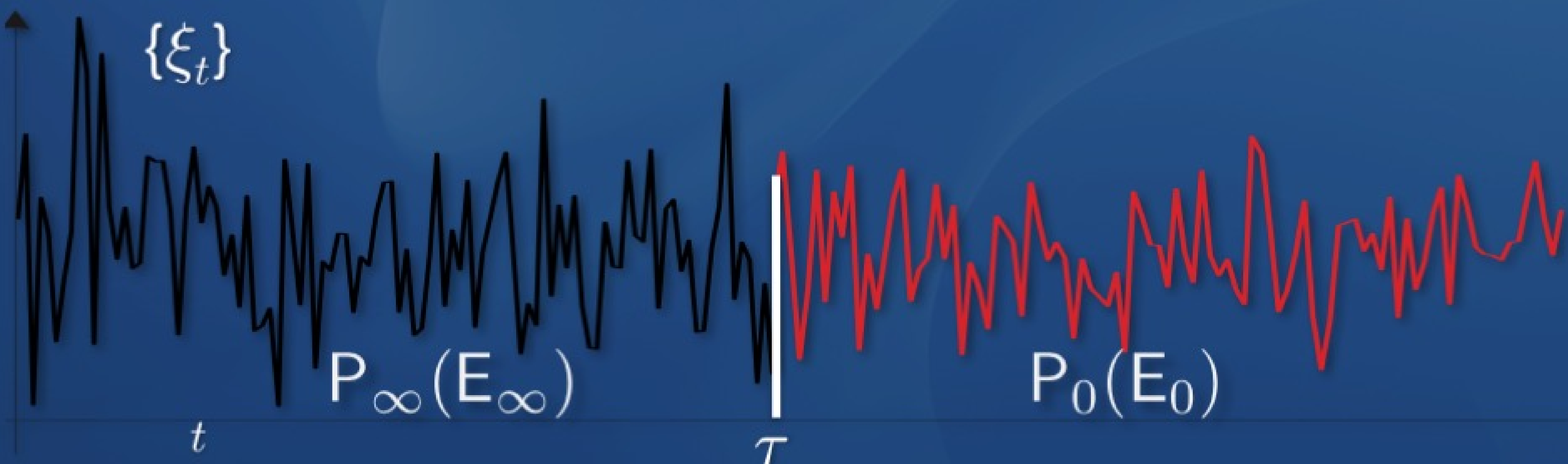
τ : first entry time controlled by $\{X_t\}$, **I want to detect it**

$\{X_t\}$ observable and $\mathcal{A}$ known: trivial.

$\{X_t\}$ (partially) hidden and/or $\mathcal{A}$ unknown: challenging.

We observe process $\{\xi_t\}$.





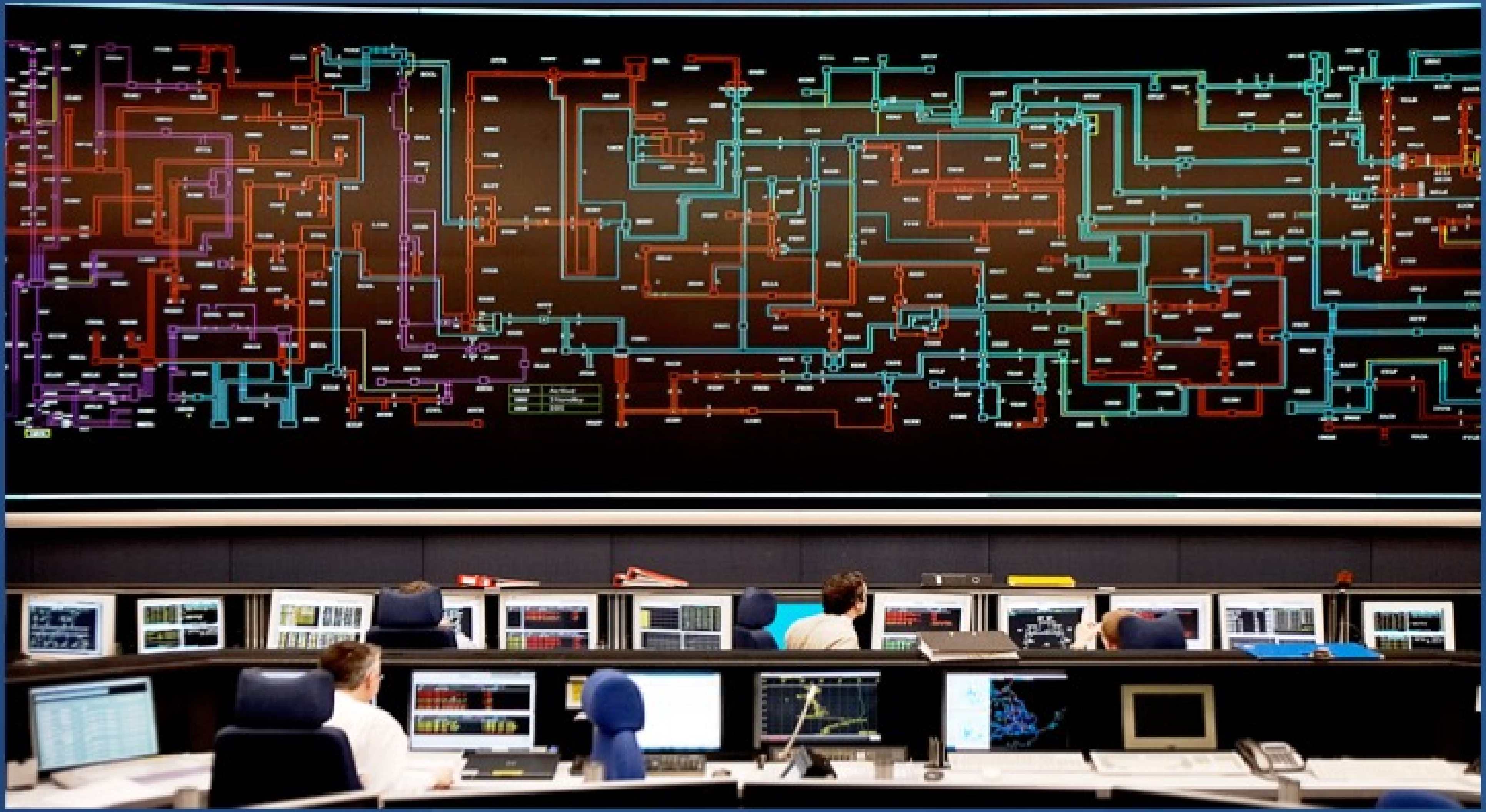
Sequential Change Detection Problem

- First-entry: **Model for change-imposing mechanism.**
- Unifies existing formulations
- Helps understanding of existing metrics
- May lead to new formulations and better detectors.

Goal: detect occurrence of τ

τ is a first entry time controlled by the process $\{X_t\}$.

T is a stopping time adapted to the filtration generated by the observation sequence $\{\xi_t\}$.



Immediate Detection

Power Grid:

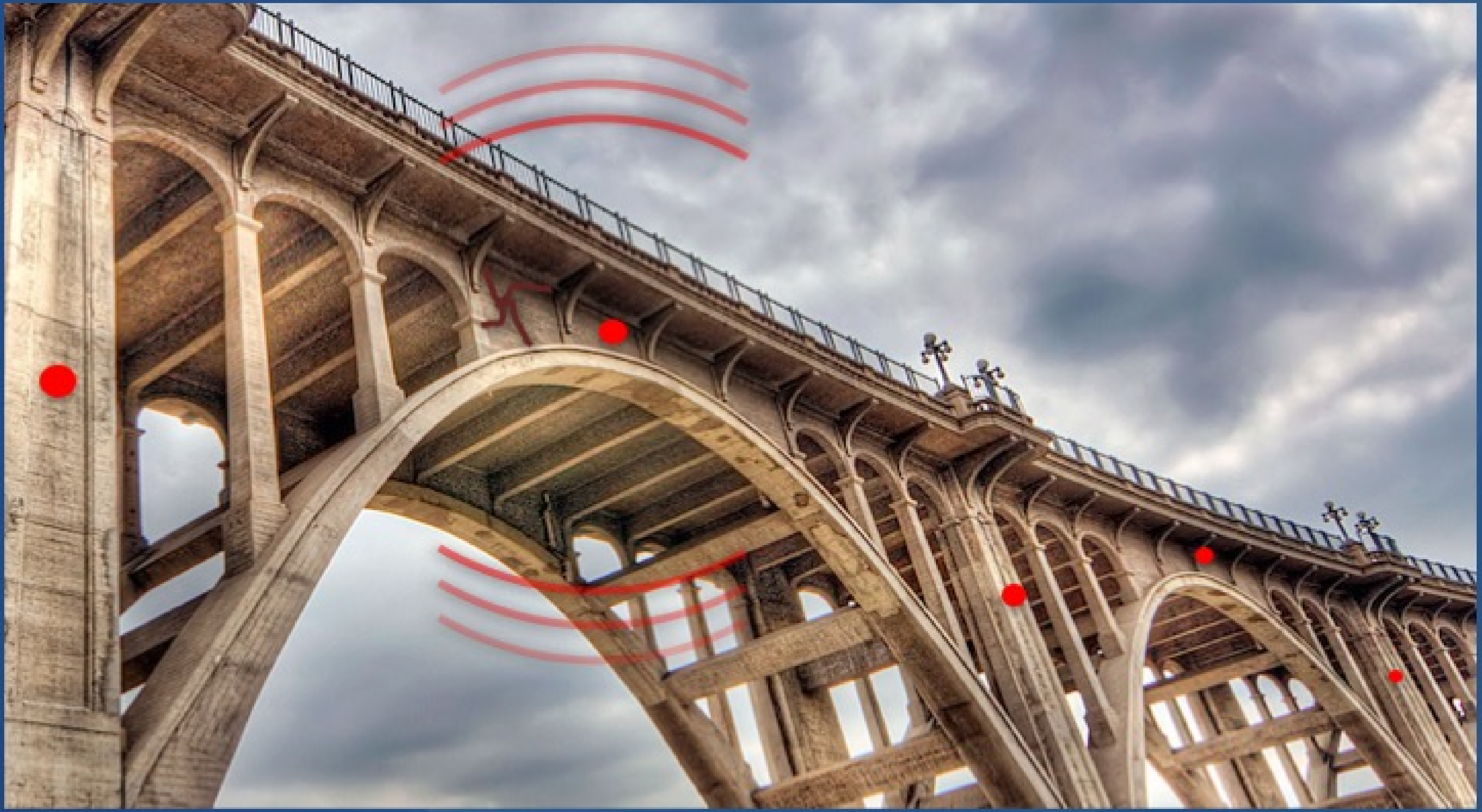
Dependent

X_t: Energy at major points in the grid.

$\xi_t = X_t + W_t$ noisy measurements.

$\mathcal{A}$: If $X_t \in \mathcal{A}$ then, after short time major blackout.

$\mathcal{A}$ is known



Delayed Detection

Structural health monitoring:

Dependent

X_t: Vibrations at every point of the structure (state)

$\xi_t = \mathbf{A}X_t + W_t$: Noisy measurements

$\mathcal{A}$: If $X_t \in \mathcal{A}$ then cracks (change in the structure)

$\mathcal{A}$ known or unknown.

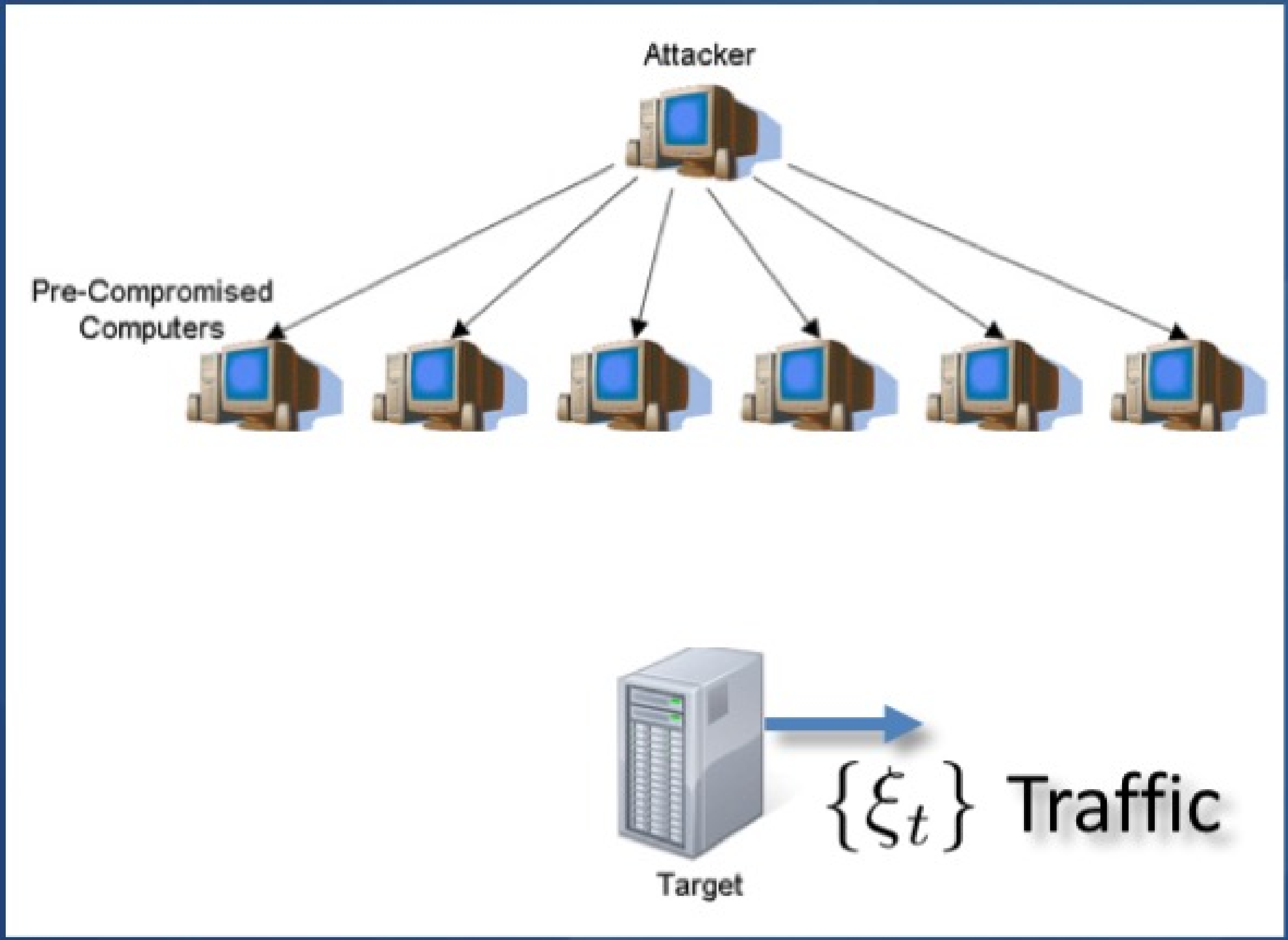
Independent $\{X_t\}$ and $\{\xi_t\}$?:

X_t: Field coordinates of the ball

ξ_t: Noisy vibration measurements

Independent

$\mathcal{A}$: Volume under the goal net.



At some point in time:
Attack!!!

Attacker no access to observations…

$\{X_t\}, \{\xi_t\}$ independent

There are important applications where the two process are independent.

However, in the majority of cases the two process **are dependent!!!**

A more general model

In the bridge example we argued that the change is imposed by a first entry mechanism:

$$\tau = \inf\{t > 0 : \; X_t \in \; \mathcal{A}\}$$

But we can have something far more complicated:

$$\tau = \inf\big\{t > 0 : \; \{X_{t-M}, \ldots, X_t\} \in \mathcal{A}_M\big\}$$

More general model than first entry, for change imposing mechanism:

τ : Stopping time adapted to history of $\{X_t\}$

τ is a stopping time controlled by the process $\{X_t\}$.

T is a stopping time controlled by the observation sequence $\{\xi_t\}$.

If stopping rule for τ known, then we should use it!

In example $\tau = \inf\Big\{t > 0 : \ \{X_{t-M}, \ldots, X_t\} \in \mathcal{A}_M\Big\}$

instead of guessing ???

safer to consider unknown stopping rule $\tau =$???

We assume that we know:

- τ : adapted to the history $\{X_t\}$
- $f_t(X_t, \xi_t | X_{t-1}, \xi_{t-1}, \ldots, X_1, \xi_1)$

Performance measures

Known change imposing mechanism

Delayed detection

$$\inf_T \mathsf{E}[T-\tau|T>\tau]$$
$$\text{subject to}: \mathsf{P}_\infty(T\le\tau)\le\alpha$$

Shiryaev (1961)

Hard limited detection delay

$$\sup_T \mathsf{P}(T\le\tau+M|T>\tau)$$
$$\text{subject to}: \mathsf{P}_\infty(T\le\tau)\le\alpha$$

Immediate detection

$$\sup_T \mathsf{P}_\infty(T=\tau|T\ge\tau)$$
$$\text{subject to}: \mathsf{P}_\infty(T<\tau)\le\alpha$$

Unknown change imposing mechanism

$$\inf_T \mathcal{J}(T) = \inf_T \sup_\tau \mathsf{E}[T-\tau|T>\tau]$$

$$\text{subject to}: \mathsf{E}_\infty[T] \geq \gamma$$

Worst-case analysis

Independent $\{X_t\}$ and $\{\xi_t\}$

$$\inf_T \sup_{t>0} \mathsf{E}_t[T-t|T>t]$$

$$\text{subject to}: \mathsf{E}_\infty[T] \geq \gamma$$

Pollak (1985)

Worst-case scenario over $\{X_t\}$ **NOT** $\{\xi_t\}$

Dependent $\{X_t\}$ and $\{\xi_t\}$

$$\inf_T \sup_{t>0} \operatorname{ess\,sup} \mathsf{E}_t[T-t|T>t, X_1, \ldots, X_t]$$

$$\text{subject to}: \mathsf{E}_\infty[T] \geq \gamma$$

Lorden?? (1971)

$$\sup_T \mathcal{P}(T) = \sup_T \inf_\tau \mathsf{P}(T \le \tau + M | T > \tau)$$

$$\text{subject to : } \mathsf{E}_\infty[T] \ge \gamma$$

Independent $\{X_t\}$ and $\{\xi_t\}$

$$\sup_T \inf_{t>0} \mathsf{P}_t(T \le t + M | T > t)$$
$$\text{subject to : } \mathsf{E}_\infty[T] \ge \gamma$$

Pollak like

Dependent $\{X_t\}$ and $\{\xi_t\}$

Lorden like

$$\sup_T \inf_{t>0} \operatorname{ess\,inf} \mathsf{P}_t(T \le t + M | T > t, X_1, \ldots, X_t)$$
$$\text{subject to : } \mathsf{E}_\infty[T] \ge \gamma$$

Hidden Markov Model (Fuh, Mei, Tartakovsky)

$\{(z_t,\xi_t)\}:\ \{z_t\}$ HMP, $\{\xi_t\}$ Observations

$$f_i(z_t,\xi_t|z_{t-1},\xi_{t-1},\ldots,z_1,\xi_1)=g_i(z_t|z_{t-1})h_i(\xi_t|z_t)$$

$$i=\infty,0$$

By considering only observations, resulting pdfs are change-time dependent: No stationarity!!!

$$\text{ess sup}\,\mathsf{E}_t[T-t|T>t,\xi_1,\ldots,\xi_t]$$

Conditioning on the pair process we obtain **stationary** conditional pdfs.

$$\text{ess sup}\,\mathsf{E}_t[T-t|T>t,\xi_1,z_1,\ldots,\xi_t,z_t]$$

Change mechanism consults $\{z_t\}$ **AND** $\{\xi_t\}$

Examples

Immediate detection

$$\tau = \inf\{t > 0 :\ X_t \in \ \mathcal{A}\}$$

Known : $\mathcal{A},\ f_\infty(X_t, \xi_t | X_{t-1}, \xi_{t-1}, \ldots, X_1, \xi_1)$

$$\sup_T \mathsf{P}_\infty(T = \tau | T \geq \tau)$$

$$\text{subject to : } \mathsf{P}_\infty(T < \tau) \leq \alpha$$

Define $\varpi_t = \mathsf{P}_\infty(\tau = t | \xi_1, \ldots, \xi_t)$

For i.i.d. pair process $\{(X_t, \xi_t)\}$

$$\varpi_t = \pi_t \prod_{k=0}^{t-1}(1-\pi_k), \quad \text{where } \pi_t = \mathsf{P}_\infty(X_t \in \mathcal{A}|\xi_t)$$

Optimum stopping time

$$T_o = \inf\{t > 0 : \pi_t \geq \nu\}, \quad \nu \in (0,1)$$

Threshold selected to satisfy constraint with equality.

For a state-space Gaussian linear model

$$X_t = \mathbf{A}X_{t-1} + W_t$$

$$\xi_t = B'X_t + v_t$$

Assume change rare

$$\varpi_t \approx \pi_t = \mathsf{P}(X_t \in \mathcal{A}|\xi_1, \ldots, \xi_t)$$

$$X_t \sim \mathcal{N}(\hat{X}_{t|t}, \Sigma_{t|t})$$

Kalman Filter

$$T_o = \inf\{t > 0, \pi_t \geq \nu\}$$

Hard Limited Delay: $\mathsf{P}(T \le \tau + M | T > \tau)$

Only for $M = 1$: $\mathsf{P}(T = \tau + 1 | T > \tau)$

Detection with the **first** sample under alternative regime

$$\mathcal{P}_{\mathsf{S}}(T) = \mathsf{P}(T = \tau + 1 | T > \tau)$$ Shiryaev like

$$\mathcal{P}_{\mathsf{P}}(T) = \inf_{t>0} \mathsf{P}_t(T = t + 1 | T > t)$$ Pollak like

$\{X_t\}$, $\{\xi_t\}$ independent

$$\mathcal{P}_{\mathsf{L}}(T) = \inf_{t>0} \operatorname{ess\,inf} \mathsf{P}_t(T = t + 1 | T > t, \xi_1, \ldots, \xi_t)$$

$\{X_t = \xi_t\}$

Lorden like

$$\sup_T \mathcal{P}_{\mathsf{S}}(T)$$
s.t. $\mathsf{P}_\infty(T \le \tau) \le \alpha$

$$\sup_T \mathcal{P}_{\mathsf{P(L)}}(T)$$
s.t. $\mathsf{E}_\infty[T] \ge \gamma$

$$T_{\mathsf{Sh}} = \inf \left\{ t > 0 : \frac{f_0(\xi_t)}{f_\infty(\xi_t)} \geq \nu \right\}$$

Shewhart test (1931)

Optimality: Bojdecki (1979): Shiryaev like

Pollak and Krieger (2013): Pollak like

Moustakides (2014): Lorden like

Pollak and Krieger (2013): Multiple post-change possibilities.

Moustakides (2014): Post change time variation

Markovian observations

$$\mathcal{P}_{\mathsf{L}}(T) = \inf_{t>0} \operatorname{ess\,inf} \mathsf{P}_t(T = t+1 | T > t, \xi_1, \ldots, \xi_t)$$

$$\sup_T \mathcal{P}_{\mathsf{L}}(T), \quad \text{subject to : } \mathsf{E}_\infty[T] \geq \gamma$$

Markovian pre- and post-change observations $\{\xi_t\}$

$$T_{\mathsf{Sh}} = \inf \left\{ t > 0 : c(\xi_{t-1}) \frac{f_0(\xi_t | \xi_{t-1})}{f_\infty(\xi_t | \xi_{t-1})} \geq \nu(\xi_t) \right\}$$

Applies **only** to the Lorden-like measure

Denote conditional LR : $\mathsf{L}(\xi_1, \xi_0) = \frac{f_0(\xi_1|\xi_0)}{f_\infty(\xi_1|\xi_0)}$

Define $c(\xi) > 0,\ \nu(\xi) > 1,$ through equations :

$$\mathsf{P}_0\big(c(\xi_0)\mathsf{L}(\xi_1, \xi_0) \geq \nu(\xi_1)|\xi_0\big) = \beta \in (0,1),\ \forall \xi_0$$

$c(\xi), \nu(\xi)$ depend on β

Forces test to be equalizer

$$\nu(\xi_0) = 1 + \mathsf{E}_\infty\left[\nu(\xi_1)\mathbb{1}_{\{c(\xi_0)\mathsf{L}(\xi_1,\xi_0)<\nu(\xi_1)\}}|\xi_0\right]$$

$\nu(\xi_0) = \mathsf{E}_\infty[T_{\mathsf{Sh}}|\xi_0]$

If ξ_0 pre-change with pdf $g_\infty(\xi)$ enforce FA equality :

$$\mathsf{E}_\infty[T_{\mathsf{Sh}}] = \mathsf{E}_\infty\big[\mathsf{E}_\infty[T_{\mathsf{Sh}}|\xi_0]\big]$$

$$= \int \nu(\xi_0) g_\infty(\xi_0)\, d\xi_0 = \gamma.$$

Functions $c(\xi), \nu(\xi)$ and detection probability β can be computed numerically

$$\mathsf{P}_\infty : \quad \xi_t = w_t, \qquad w_t \sim \mathcal{N}(0,1) \text{ i.i.d.}$$

$$\mathsf{P}_0 : \quad \xi_t = 0.5\xi_{t-1} + w_t$$

False alarm constraint : $\gamma = 100$

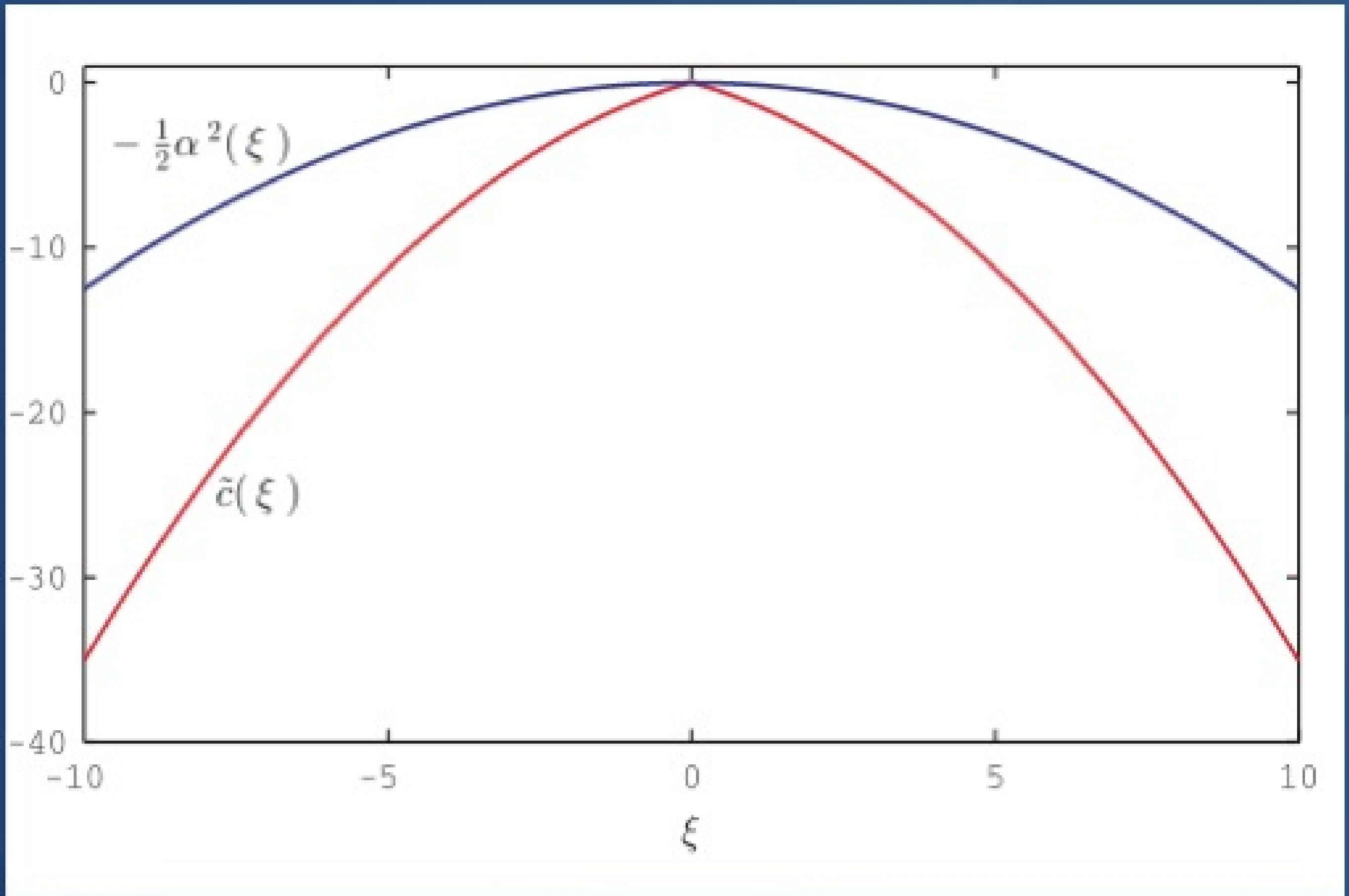
Optimum (after taking the logarithm):

$$T_{\mathsf{Sh}} = \inf\{t > 0 : \tilde{c}(\xi_{t-1}) + 0.5\xi_{t-1}\xi_t \geq \tilde{\nu}(\xi_t)\}$$

$$\tilde{c}(\xi) = \log c(\xi) - 0.125\xi^2, \quad \tilde{\nu}(\xi) = \log \nu(\xi)$$

Naïve - Compare conditional LR to constant threshold:

$$\mathcal{T} = \inf\{t > 0 : -0.125\xi_{t-1}^2 + 0.5\xi_{t-1}\xi_t \geq \tilde{\nu}\}$$



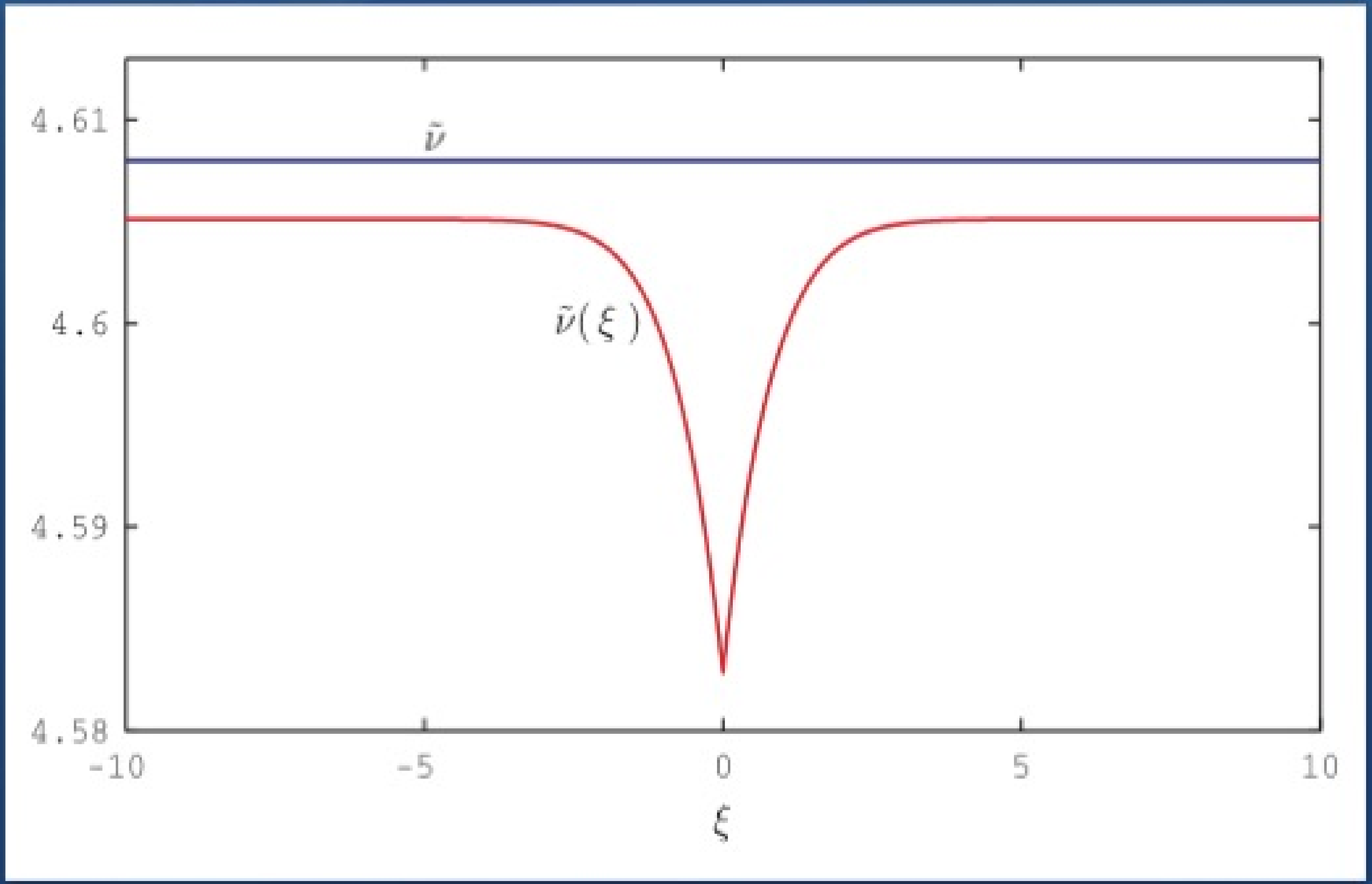
Optimum: worst-case performance: 0.022

Holy Grail: Solve Lorden's formulation

Naïve: worst-case performance: 0.0

$$-0.125\xi_{t-1}^2 + 0.5\xi_{t-1}\xi_t \geq \tilde{\nu}$$

ess inf for $\xi_{t-1} = 0: \quad 0 \not\geq \tilde{\nu}$



Acknowledgements

NSF: CIF-1513373 through Rutgers University Collaboration program with UIUC

Also partially supported by the project FEDER (XterM, University of Rouen, France

Ce projet est cofinancé par l'Union européenne.
L'Europe s'engage en Normandie
avec le Fonds européen de Développement Régional (FEDER)

RÉGION NORMANDIE

Thank you for your attention!