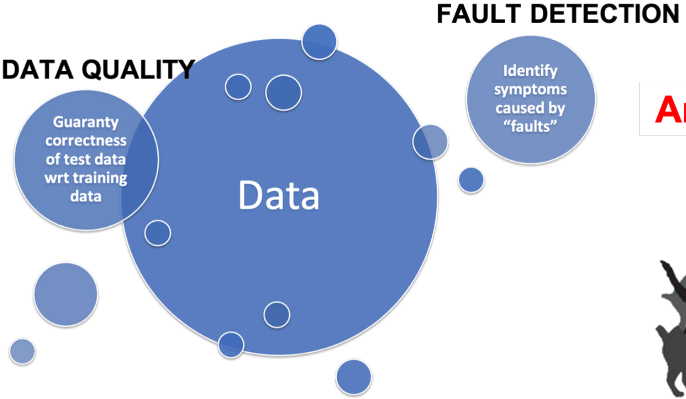
Improving anomaly detection in data streams with the Christoffel function

Louise Travé-Massuyès, Jean-Bernard Lasserre
Kevin Ducharlet

March 25, 2022

Université Fédérale
Toulouse Midi-Pyrénées

Anomaly detection

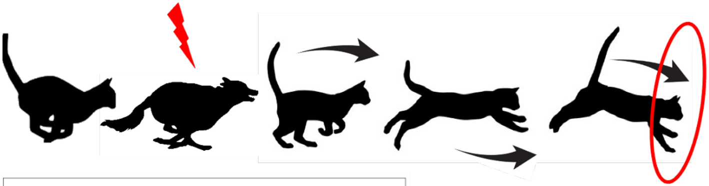


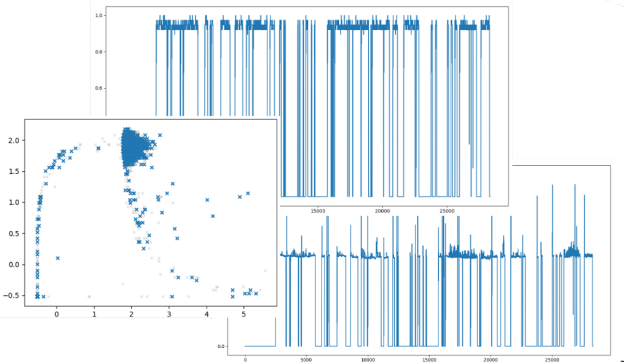
Are there intruders ?

Anomaly

- an instance from a distinct distribution
- a rare or low-probability instance

Anomaly detection in data streams





Data stream requirements

- detection on the fly
- incremental update
- low complexity
- complex distributions

This work

A **hybrid AI** anomaly detection method for data streams that:

- leverages the Christoffel function
  - related to the Christoffel-Darboux kernel borrowed from the theory of approximation and orthogonal polynomials
  - advocated for data mining by J.-B. Lasserre and E. Pauwels (2019)
- benefits from a clean algebraic framework
- fulfils all data stream requirements
- needs little tuning

This work

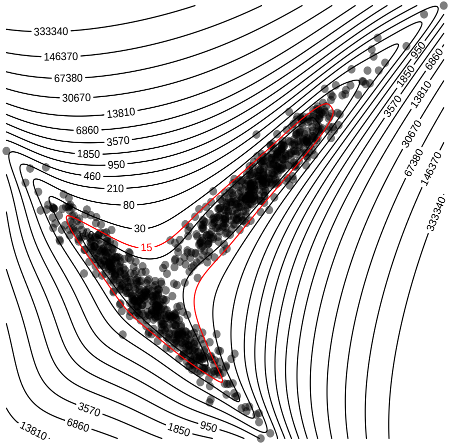
A **hybrid AI** anomaly detection method for data streams that:

- leverages the Christoffel function
  - related to the Christoffel-Darboux kernel borrowed from the theory of approximation and orthogonal polynomials
  - advocated for data mining by J.-B. Lasserre and E. Pauwels (2019)
- benefits from a clean algebraic framework
- fulfils all data stream requirements
- needs little tuning

A collaboration between two ANITI chairs:

- **Polynomial Optimization** for Machine Learning and Data Analysis (*Jean-Bernard Lasserre*)
- Synergistic Transformations in Model Based and Data Based **Diagnosis** (*Louise Travé-Massuyès*)

Capturing the shape of a cloud of data points



Consider a cloud of data points $(x(i))_{i \in \mathbb{N}} \subset \mathbb{R}^p$

The red curve is the level set:
$\mathcal{L}_\gamma = \{\mathbf{x} : Q_d(\mathbf{x}) \leq \gamma\}, \gamma \in \mathbb{R}_+$

of a certain polynomial $Q_d \in \mathbb{R}[x1, x2]$ of degree $2d$.

Notice that $\mathcal{L}_\gamma$ captures the shape of the cloud.

The Christoffel function

- Let μ be a Borel measure on a compact set $\Omega \subset \mathbb{R}^p$ with nonempty interior,
- Form the vector $\mathbf{v}_d(\mathbf{x})$ from a basis of p-variate polynomials of degree at most d:

$$\mathbf{v}_d(\mathbf{x}) = (P_1(\mathbf{x}), \ldots, P_{s(d)}(\mathbf{x}))^T \quad \text{of size } s(d) = \binom{p+d}{p}.$$

$$Q_d^\mu(\mathbf{x}) = \mathbf{v}_d(\mathbf{x})^T \underbrace{\mathbf{M}_d(\mu)^{-1}}_{\text{Moment matrix of } \mu} \mathbf{v}_d(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^p$$

The **Christoffel function** $\Lambda_d^\mu : \mathbb{R}^p \to \mathbb{R}_+$ is defined by:

$$\Lambda_d^\mu(\mathbf{x})^{-1} = Q_d^\mu(\mathbf{x})$$

Λ_d^μ encodes properties of the underlying measure μ.

Empirical measure

In our case

$$\mu_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{x(i)}$$

is the EMPIRICAL measure associated with the cloud of data points $(x(i))_{i \leq n}$ sampled from an unknown measure μ on Ω.

☞ ... and quite remarkably

The level sets of $\Lambda_d^{\mu_n}(\mathbf{x})^{-1}$ match the density variations of the cloud of points $(x(i))_{i \leq n}$

$\rightarrow \Lambda_d^{\mu_n}(\mathbf{x})^{-1}$ is a good scoring function for anomaly detection

In particular, the level set

$$\{\mathbf{x} \in \mathbb{R}^p \, : \, \Lambda_d^{\mu_n}(\mathbf{x})^{-1} \leq \binom{p+d}{p}\}$$

identifies the support Ω of μ, even for moderate values of d.

Dealing with a data stream

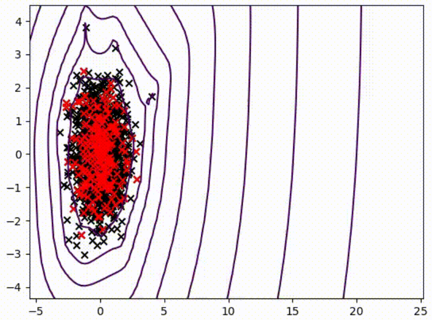
Incremental update of $\Lambda_d^{\mu_n}(\mathbf{x})^{-1}$ with rank-one update of the inverse $\mathbf{M}_d(\mu_n)^{-1}$

When a point ξ is added to the cloud of n points, i.e.,

$$\mu_n \rightarrow \frac{1}{n+1}\,(n\,\mu_n + \delta_\xi)$$

$\rightarrow$ a new cloud with n+1 points

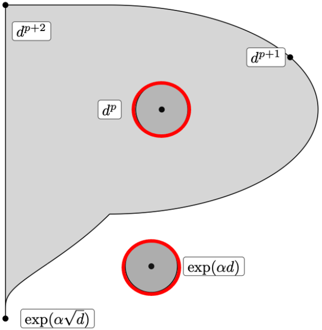
☞ The Sherman-Morrisson-Woodbury formula allows for a simple RANK-ONE UPDATE of the inverse $\mathbf{M}_d(\mu_n)^{-1}$

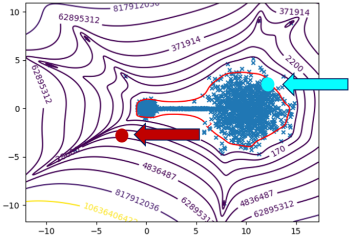


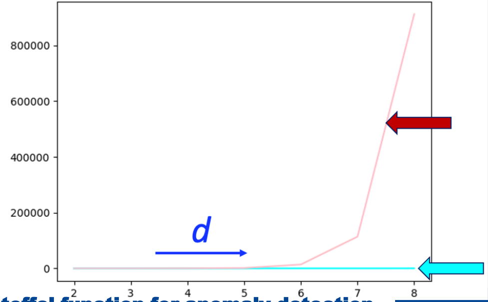
Crucial property: growth rate of the Christoffel function

As d grows, $\Lambda_d^{\mu}(\mathbf{x})^{-1}$ has:

$$\begin{cases} \text{POLYNOMIAL growth} & \text{INSIDE } \Omega \\ \text{EXPONENTIAL growth} & \text{OUTSIDE } \Omega \end{cases}$$







ANITI

Université Fédérale Toulouse Midi-Pyrénées

From scoring to binary anomaly prediction

The Dynamic Christoffel Growth Method: Dy-CG

$$\begin{array}{rcl} \mu_n & : & n \quad \text{first data points of the data stream.} \\ \Delta = (d_j)_{1 \leq j \leq k} & : & \text{sequence of} \quad k \quad \text{degrees in ascending order.} \end{array}$$

For every incoming $x(n+i)_{i \geq 1}$ of the data stream

- Identify an anomaly candidate with a reference score:

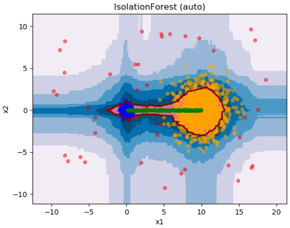
$$d_{ref} = d_k = sup(\Delta) \quad \longrightarrow \quad \Lambda_{d_k}^{\mu_{n+i-1}}(x(n+i))^{-1}$$

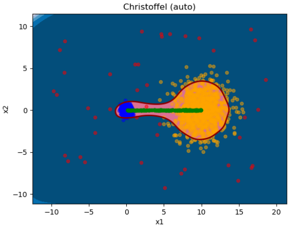
- Refute/confirm anomaly from the exponential growth of the scores:

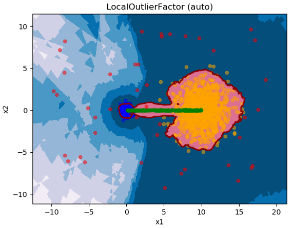
$$\begin{array}{ccc} d_1 & \longrightarrow & \Lambda_{d_1}^{\mu_{n+i-1}}(x(n+i))^{-1} \\ \vdots & & \vdots \\ d_k & \longrightarrow & \Lambda_{d_k}^{\mu_{n+i-1}}(x(n+i))^{-1} \end{array}$$

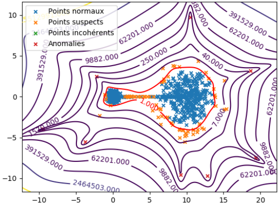
Some comparisons

Isolation Forest, Simple Christoffel, Local Outlier Factor







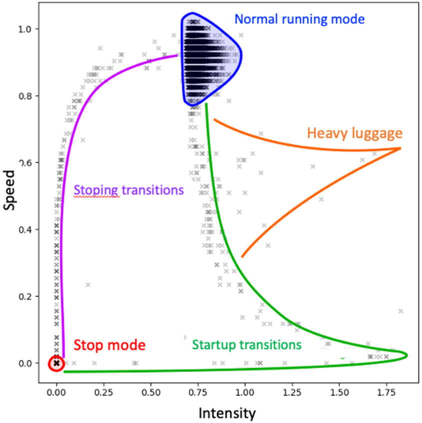


Our method: *Dy-CG*

Experiments on an industrial luggage conveyor data

Carl Berger-Levrault project - Multi-mode system, highly non linear data

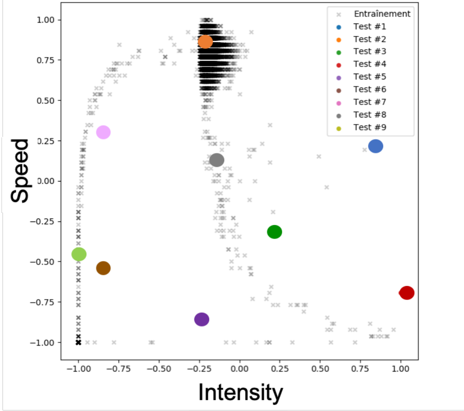
Cifre thesis of Kevin Ducharlet

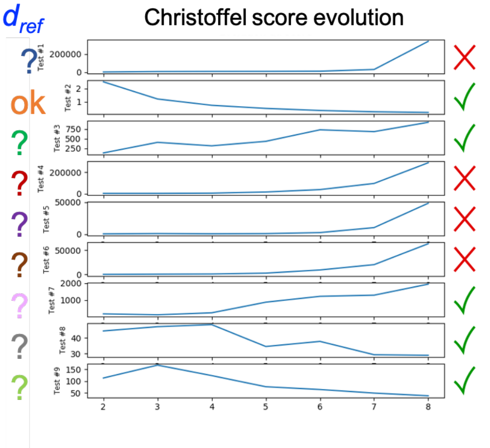


Data : conveyor speed and motor intensity labelled by modes

Experiments on an industrial luggage conveyor data

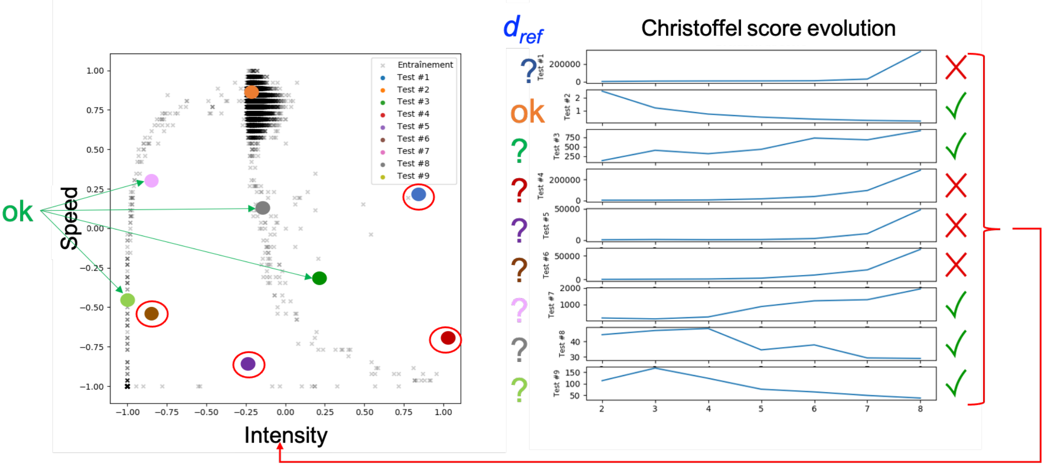
Anomaly candidates from score for d_{ref}, refutal/confirmation from exponential growth as a function of d





Experiments on an industrial luggage conveyor data

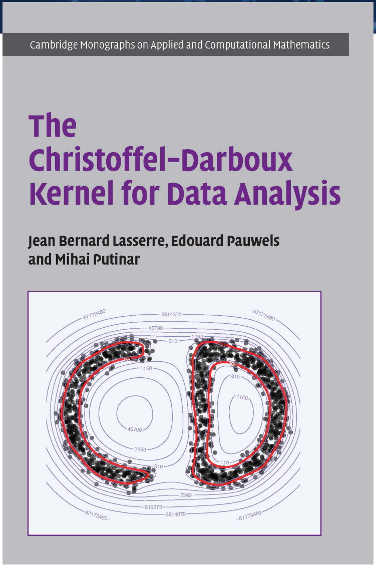
Anomaly candidates from score for d_{ref}, refutal/confirmation from exponential growth as a function of d



ANITI Université Fédérale Toulouse Midi-Pyrénées

Conclusion

- *Dy-CG* is a simple and easy-to-use method with little tuning that achieves excellent results compared to other more tricky anomaly detection methods
- The Christoffel-based scoring function is directly issued from the moments of the measure underlying the set of data points
- It leverages gthe rowth rate properties of the Christoffel function
- It nicely deals with data streams thanks to incremental update
- On going work:
  - adding forgetting ability
  - scaling up to high dimensions
  - designing metrics to evaluate non supervised anomaly detection methods


Cambridge Monographs on Applied and Computational Mathematics
The Christoffel–Darboux Kernel for Data Analysis
Jean Bernard Lasserre, Edouard Pauwels and Mihai Putinar