B Ravi Kiran,  
Université Lille 3, CRISTaL  
Joint work with Mathieu Andreux  

beedotkiran@gmail.com  

June 20, 2017
Overview

1. Anomaly detection

2. Time series representation

3. Streaming Subspace Tracking

4. Results
Motivation: Anomaly detection

- Areas: Industrial processes, medical and satellite telemetry, Finance
- Anomaly Detection: $x(t)$, $t$ where signal “deviates” from the local mean value.
- $x(t)$ are observations over time $t$ where new data arrives over time $t$.
- High volumes of data are generated per day (in the GBs).

Requirements:

- Time series representation is robust to variation in scale of pseudo-periodicities (window size).
- Streaming time series anomaly detection to handle large abouts of data.

Require an online multiscale anomaly detection algorithm.
Yahoo! and Numenta Datasets

Yahoo! unsupervised anomaly detection Benchmark [8] provides datasets with annotated anomalies and changepoints.

Numenta Anomaly Benchmark (NAB) [9] provides an evaluation of streaming time series anomaly detection algorithms.

The datasets contain various types of anomalies: level shifts/change-points, point anomalies, change in periodicities, value drifts, change in envelopes, linear trends.
Anomaly detection problem

- **Formulation:**
  - Track the principal direction given a scale/lag $p$ for the design matrix of time series.
  - Evaluate the reconstruction error to measure deviation from the rest of the windows.
  - Evaluate across multiple lags ($p$)
  - Characterize anomalies by their variation in reconstruction error across scale of lag-window size.

**Related work:**
- Streaming anomaly detection by subspace tracking [5]
- Tracking correlations over multi-scale windows for frequent motif extraction [12]
- Multi-scale anomaly detection offline [3]
We build a lag matrix over a window of size $p$

$$X_t^p = [x_t, x_{t-1}, \ldots, x_{t-p+1}]^T \in \mathbb{R}^p$$
Multiscale Lagmatrix

\[ X_t^p = [x_t, x_{t-1}, \ldots, x_{t-p+1}]^T \in \mathbb{R}^p \]
Principal Subspace Tracking

Streaming PCA:
- Dimensionality reduction for time series lag embedding
- Recursive update for principal subspace

Linear Principal Component Analysis criterion:

$$J(w_t) = E \left[ \| X_t - w_t w_t^T X_t \|^2 \right]$$

$w_t \in \mathbb{R}^{p \times r}$ At the global minimum for $w_t$ shall contain the $r$ dominant eigen-vectors.

- Online principal subspace tracking of the lagmatrix to track correlations: SPIRIT algorithm [11]
- Given $X^p \in \mathbb{R}^{T \times p}$, $w_p$ is defined as the 1-D projection capturing most of the energy of the data samples:

$$w_p = \arg \min_{\|w\|=1} \sum_{t=1}^{T} \| X_t^p - (w_p w_p^T) X_t^p \|^2$$
Algorithm 1 Streaming PCA

Initialization: \( w_j \leftarrow 0 \), \( \sigma_j^2 \leftarrow \epsilon \) with \( \epsilon \ll 1 \)

for \( t = 1, \ldots, T \) do
    for \( j = 1, \ldots, J \) do
        \( Z^j_t \leftarrow H^{T_j} X^j_t \)
        \( y^j_t \leftarrow w^j_t Z^j_t \)
        \( \sigma_j^2 \leftarrow \sigma_j^2 + (y^j_t)^2 \)
        \( e^j_t \leftarrow Z^j_t - y^j_t w_j^j \)
        \( w_j \leftarrow w_j + \sigma_j^{-2} y^j_t e^j_t \)
        \( \pi^j_t \leftarrow w^T_j Z^j_t \)
        \( \tilde{Z}^j_t \leftarrow \pi^j_t w_j^j \)
        \( \alpha^j_t \leftarrow \| \tilde{Z}^j_t - Z^j_t \|^2 \)
    end for
end for

return \( \alpha \in \mathbb{R}^{T \times J} \)

Given \( x(t) \in \mathbb{R} \) for \( t = 1 : T \)

- We evaluate the lag-matrix \( X \in \mathbb{R}^{T \times p} \) where \( p = 2^j \).
- For each vector \( X_t \in \mathbb{R}^p \) we perform a change of basis \( Z_t := \Phi^T X_t \)
- We require a unitary transform to
  - localize a deviation from the local mean and variations.
  - Preserve the variance.
- Haar transform \( \Phi = H : \)
  \[
  H_{2N} = \frac{1}{\sqrt{2}} \begin{bmatrix} H_N \otimes [1, 1] \\
  I_N \otimes [1, -1] \end{bmatrix}
  \]

(CRISTaL)
Streaming PCA point cloud (2d embedding only)

Centered time series and anomalies

Signal
is_anomaly
Error Spectrogram

Reconstruction error of the lag-matrix calculated in logarithmic scales:
When passing from one scale $p_j$ to the next $p_{j+1} = 2p_j$, instead of rebuilding a lag matrix $X_t^{j+1}$ whose size doubles, it builds a reduced lag matrix $Z_t^{j+1}$ by considering the projection of each component of size $p_j$ on the principal direction obtained at this scale, i.e.

$$Z_t^{j+1} = [w_j^T Z_t^j, w_j^T Z_{t-2j}]^T$$

with $Z_t^1 = X_t^1$.

The principal direction at scale $p_{j+1}$ is then obtained by applying the streaming PCA algorithm on this reduced representation.

**Figure**: Hierarchical PCA.
Aggregating Multi-scale Anomaly score

At time $t$, we denote by $\tilde{X}_t^p$ the projection of $X_t^p$ upon $w^p$ (at this time step), i.e. $\tilde{X}_t^p = w_p^T X_t^p$. We obtain $\alpha_t \in \mathbb{R}^{T \times J}$, we propose the following ways to aggregated the $J$ scales:

1. $\|\alpha_t\|^2$: Norm of multiscale anomaly score
2. $\|\tilde{\alpha}_t - \alpha_t\|^2$: Streaming reconstruction error on anomaly score, obtained via a 2nd iteration of the streaming PCA algorithm on the multiscale anomaly score instead of the lag-matrix.
3. $\alpha_{j^*}^t$ where $j^* = \arg \min_j \sum_i (\alpha^T \alpha)_{ji}$: the anomaly score corresponding to the scale which is least correlated with others.

Performance Evaluation:

- Area under the receiver operators characteristics curve (AUC)
- Integrating the curve of the False positive rate (FPR) vs the True positive rate (TPR) obtained for all possible thresholds.
- 0 (worst value) and 1 (perfect detector)
Global-flow

- Representation $\Phi^T X_t$: Localize the anomaly in a basis
- Multiscale Anomaly Score: Compose anomaly scores
### Results

#### Multi-scale score-Norm $\|\alpha_t\|^2$ (PC=1)

<table>
<thead>
<tr>
<th>Method / AUCs</th>
<th>Bench 1</th>
<th>Bench 2</th>
<th>Bench 3</th>
<th>Bench 4</th>
<th>NAB</th>
</tr>
</thead>
<tbody>
<tr>
<td>fixed-scale</td>
<td>0.828 ± 0.240</td>
<td>0.835 ± 0.180</td>
<td>0.614 ± 0.108</td>
<td>0.568 ± 0.160</td>
<td>0.815 ± 0.238</td>
</tr>
<tr>
<td>fixed-scale-haar</td>
<td>0.826 ± 0.238</td>
<td>0.878 ± 0.143</td>
<td>0.617 ± 0.115</td>
<td>0.576 ± 0.157</td>
<td>0.812 ± 0.232</td>
</tr>
<tr>
<td>multiscale-lagmatrix</td>
<td>0.884 ± 0.232</td>
<td>0.978 ± 0.057</td>
<td>0.816 ± 0.092</td>
<td>0.696 ± 0.157</td>
<td>0.879 ± 0.199</td>
</tr>
<tr>
<td>hierarchical-approx</td>
<td>0.871 ± 0.236</td>
<td>0.997 ± 0.002</td>
<td>0.980 ± 0.025</td>
<td>0.897 ± 0.104</td>
<td>0.900 ± 0.189</td>
</tr>
<tr>
<td>multiscale-haar</td>
<td>0.906 ± 0.231</td>
<td>0.989 ± 0.019</td>
<td>0.992 ± 0.019</td>
<td>0.892 ± 0.126</td>
<td>0.892 ± 0.198</td>
</tr>
</tbody>
</table>

#### PCA on multi-scale score $\|\tilde{\alpha}_t - \alpha_t\|^2$ (PC=1)

<table>
<thead>
<tr>
<th>Method / AUCs</th>
<th>Bench 1</th>
<th>Bench 2</th>
<th>Bench 3</th>
<th>Bench 4</th>
<th>NAB</th>
</tr>
</thead>
<tbody>
<tr>
<td>fixed-scale</td>
<td>0.632 ± 0.264</td>
<td>0.754 ± 0.206</td>
<td>0.533 ± 0.124</td>
<td>0.525 ± 0.133</td>
<td>0.700 ± 0.247</td>
</tr>
<tr>
<td>fixed-scale-haar</td>
<td>0.649 ± 0.251</td>
<td>0.723 ± 0.194</td>
<td>0.514 ± 0.110</td>
<td>0.522 ± 0.129</td>
<td>0.699 ± 0.244</td>
</tr>
<tr>
<td>multiscale-lagmatrix</td>
<td>0.895 ± 0.218</td>
<td>0.997 ± 0.006</td>
<td>0.993 ± 0.017</td>
<td>0.959 ± 0.063</td>
<td>0.891 ± 0.194</td>
</tr>
<tr>
<td>hierarchical-approx</td>
<td>0.859 ± 0.233</td>
<td>0.997 ± 0.002</td>
<td>0.961 ± 0.071</td>
<td>0.895 ± 0.108</td>
<td>0.884 ± 0.204</td>
</tr>
<tr>
<td>multiscale-haar</td>
<td>0.888 ± 0.219</td>
<td>0.988 ± 0.031</td>
<td>0.956 ± 0.059</td>
<td>0.898 ± 0.106</td>
<td>0.886 ± 0.178</td>
</tr>
</tbody>
</table>

#### Least correlated scale $\alpha^*_t$ where $j^* = \arg\min_j \sum_i (\alpha^T \alpha)_{ji}$ (PC=1)

<table>
<thead>
<tr>
<th>Method / AUCs</th>
<th>Bench 1</th>
<th>Bench 2</th>
<th>Bench 3</th>
<th>Bench 4</th>
<th>NAB</th>
</tr>
</thead>
<tbody>
<tr>
<td>fixed-scale</td>
<td>0.828 ± 0.240</td>
<td>0.835 ± 0.180</td>
<td>0.614 ± 0.108</td>
<td>0.568 ± 0.160</td>
<td>0.815 ± 0.238</td>
</tr>
<tr>
<td>fixed-scale-haar</td>
<td>0.826 ± 0.238</td>
<td>0.878 ± 0.143</td>
<td>0.617 ± 0.115</td>
<td>0.576 ± 0.157</td>
<td>0.812 ± 0.232</td>
</tr>
<tr>
<td>multiscale-lagmatrix</td>
<td>0.816 ± 0.238</td>
<td>0.773 ± 0.236</td>
<td>0.993 ± 0.017</td>
<td>0.964 ± 0.055</td>
<td>0.885 ± 0.196</td>
</tr>
<tr>
<td>hierarchical-approx</td>
<td>0.816 ± 0.238</td>
<td>0.773 ± 0.236</td>
<td>0.993 ± 0.017</td>
<td>0.964 ± 0.055</td>
<td>0.885 ± 0.196</td>
</tr>
<tr>
<td>multiscale-haar</td>
<td>0.832 ± 0.238</td>
<td>0.997 ± 0.007</td>
<td>0.799 ± 0.120</td>
<td>0.817 ± 0.123</td>
<td>0.886 ± 0.183</td>
</tr>
</tbody>
</table>
Effect of the Iterated Streaming PCA

- The error here should decorrelate the scores at different scales.
- Plotting Mean Recon. Error (Approximation) Vs. AUC (Detection)

\[
\|\alpha_t\|^2 \quad \text{Vs.} \quad \|\tilde{\alpha}_t - \alpha_t\|^2.
\]

![Graphs showing AUC score vs. average reconstruction error for Norm and Norm-PCA methods.](image)
Effect of the Iterated Streaming PCA

- The error here should decorrelate the scores at different scales.
- Plotting Mean Recon. Error (Approximation) Vs. AUC (Detection)

\[ \| \alpha_t \|^2 \quad \text{Vs.} \quad \| \tilde{\alpha}_t - \alpha_t \|^2 : \]
Effect of the Iterated Streaming PCA

- The error here should decorrelate the scores at different scales.
- Plotting Mean Recon. Error (Approximation) Vs. AUC (Detection)

\[ \| \alpha_t \|^2 \quad \text{Vs.} \quad \| \tilde{\alpha}_t - \alpha_t \|^2: \]
Effect of the Iterated Streaming PCA

- The error here should decorrelate the scores at different scales.
- Plotting Mean Recon. Error (Approximation) Vs. AUC (Detection)

\[ \| \alpha_t \|^2 \quad \text{Vs.} \quad \| \tilde{\alpha}_t - \alpha_t \|^2 \]:

![Graph showing the comparison between Norm and Norm-PCA metrics.](image-url)
Effect of the Iterated Streaming PCA

- The error here should decorrelate the scores at different scales.
- Plotting Mean Recon. Error (Approximation) Vs. AUC (Detection)

\[ \|\alpha_t\|^2 \quad \text{Vs.} \quad \|\tilde{\alpha}_t - \alpha_t\|^2: \]
Effect of Iterated Streaming PCA

- Centered and scaled Signal with anomaly
- Anomaly score
- ROC plot with AUC (anomaly)
- Covariance Matrix (scales)
Failure Cases

When the errors of reconstruction across scales remain correlated:

**Figure:** Scale correlation.

A larger scale of lag-window provides a least correlated scale.
Figure: Near zero AUC score.
Future work

Improvements on current model

- Understand the bounds on the reconstruction error $\alpha(t)$ for Streaming PCA.
- Better base-line with the multivariate zscore by calculating covariance matrix online.
- Add anomaly-score likelihood to filter the anomaly score by using a moving window gaussian neg-log score.
- Use a streaming recursively calculable multi-scale time series representation $\Phi^T X_t$: This should make use of coefficients that are calculated in the past. For now the Haar transformation $HX_t$ operates on a single vector. [4]
Future work

Other Tasks

- Anomalous time series ranking [8]
- Online Change-point evaluation [8]

Other applications:

- Unsupervised unusual action recognition in videos
- Change detection in areal/remote sensing data: hyperspectral video.
The End.
Hierarchical PCA

Initialization: $w_j \leftarrow 0$, $\sigma_j^2 \leftarrow \epsilon$ with $\epsilon \ll 1$

for $t = 1, \ldots, T$ do
  for $j = 2, \ldots, J$ do
    if $j = 1$ then
      $Z^j_t \leftarrow X^j_t$
    else
      $Z^j_t \leftarrow [\pi_t^{j-1}, (X^j_t)^T]$
    end if
    $y^j_t \leftarrow w^j_T Z^j_t$
    $\sigma_j^2 \leftarrow \sigma_j^2 + (y^j_t)^2$
    $e^j_t \leftarrow Z^j_t - y^j_t w_j$
    $w_j \leftarrow w_j + \sigma_j^{-2} y^j_t e^j_t$
    $\pi_t^j \leftarrow w^j_T Z^j_t$
    $\tilde{Z}^j_t \leftarrow \pi_t^j w_j$
    $\tilde{Z}^j_t \leftarrow \pi_t^j w_j$
    $\tilde{Z}^j_t \leftarrow \pi_t^j w_j$
    $\alpha_t^j \leftarrow \|\tilde{Z}^j_t - Z^j_t\|^2$
  end for
end for

return $\alpha_t \in \mathbb{R}^{T \times J}$
### Results PC=2

#### Multi-scale score-Norm $\|\alpha_t\|^2$ (PC=2)

<table>
<thead>
<tr>
<th>Method</th>
<th>$\alpha_t$ (PC=2)</th>
<th>$\alpha_t$ (PC=2)</th>
<th>$\alpha_t$ (PC=2)</th>
<th>$\alpha_t$ (PC=2)</th>
<th>$\alpha_t$ (PC=2)</th>
<th>$\alpha_t$ (PC=2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>fixed-scale</td>
<td>0.783 ± 0.269</td>
<td>0.918 ± 0.065</td>
<td>0.616 ± 0.142</td>
<td>0.569 ± 0.154</td>
<td>0.815 ± 0.231</td>
<td></td>
</tr>
<tr>
<td>fixed-scale-haar</td>
<td>0.808 ± 0.259</td>
<td>0.925 ± 0.074</td>
<td>0.627 ± 0.146</td>
<td>0.586 ± 0.144</td>
<td>0.811 ± 0.232</td>
<td></td>
</tr>
<tr>
<td>multiscale-lagmatrix</td>
<td>0.850 ± 0.242</td>
<td>0.969 ± 0.031</td>
<td>0.803 ± 0.116</td>
<td>0.686 ± 0.163</td>
<td>0.862 ± 0.210</td>
<td></td>
</tr>
<tr>
<td>hierarchical-approx</td>
<td>0.848 ± 0.240</td>
<td><strong>0.985 ± 0.056</strong></td>
<td><strong>0.982 ± 0.021</strong></td>
<td><strong>0.941 ± 0.079</strong></td>
<td><strong>0.876 ± 0.213</strong></td>
<td></td>
</tr>
<tr>
<td>multiscale-haar</td>
<td><strong>0.862 ± 0.245</strong></td>
<td>0.976 ± 0.021</td>
<td>0.805 ± 0.150</td>
<td>0.710 ± 0.166</td>
<td>0.873 ± 0.195</td>
<td></td>
</tr>
</tbody>
</table>

#### PCA on multi-scale score $\|\tilde{\alpha}_t - \alpha_t\|^2$ (PC=2)

<table>
<thead>
<tr>
<th>Method</th>
<th>$\alpha_t$ (PC=2)</th>
<th>$\alpha_t$ (PC=2)</th>
<th>$\alpha_t$ (PC=2)</th>
<th>$\alpha_t$ (PC=2)</th>
<th>$\alpha_t$ (PC=2)</th>
<th>$\alpha_t$ (PC=2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>fixed-scale</td>
<td>0.778 ± 0.270</td>
<td>0.908 ± 0.091</td>
<td>0.609 ± 0.133</td>
<td>0.573 ± 0.154</td>
<td>0.813 ± 0.232</td>
<td></td>
</tr>
<tr>
<td>fixed-scale-haar</td>
<td>0.804 ± 0.261</td>
<td>0.922 ± 0.079</td>
<td>0.625 ± 0.148</td>
<td>0.584 ± 0.143</td>
<td>0.811 ± 0.232</td>
<td></td>
</tr>
<tr>
<td>multiscale-lagmatrix</td>
<td>0.828 ± 0.237</td>
<td>0.872 ± 0.134</td>
<td>0.834 ± 0.172</td>
<td>0.793 ± 0.181</td>
<td>0.829 ± 0.207</td>
<td></td>
</tr>
<tr>
<td>hierarchical-approx</td>
<td><strong>0.831 ± 0.248</strong></td>
<td><strong>0.978 ± 0.084</strong></td>
<td><strong>0.976 ± 0.031</strong></td>
<td><strong>0.935 ± 0.084</strong></td>
<td><strong>0.841 ± 0.231</strong></td>
<td></td>
</tr>
<tr>
<td>multiscale-haar</td>
<td>0.816 ± 0.239</td>
<td>0.933 ± 0.088</td>
<td>0.859 ± 0.161</td>
<td>0.799 ± 0.171</td>
<td>0.807 ± 0.226</td>
<td></td>
</tr>
</tbody>
</table>

#### Least correlated scale $\alpha_t^{j*}$ where $j^* = \arg\min_j \sum_i (\alpha^T \alpha)_i$ (PC=2)

<table>
<thead>
<tr>
<th>Method</th>
<th>$\alpha_t$ (PC=2)</th>
<th>$\alpha_t$ (PC=2)</th>
<th>$\alpha_t$ (PC=2)</th>
<th>$\alpha_t$ (PC=2)</th>
<th>$\alpha_t$ (PC=2)</th>
<th>$\alpha_t$ (PC=2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>fixed-scale</td>
<td>0.783 ± 0.269</td>
<td>0.918 ± 0.065</td>
<td>0.616 ± 0.142</td>
<td>0.569 ± 0.154</td>
<td>0.815 ± 0.231</td>
<td></td>
</tr>
<tr>
<td>fixed-scale-haar</td>
<td><strong>0.808 ± 0.259</strong></td>
<td><strong>0.925 ± 0.074</strong></td>
<td><strong>0.627 ± 0.146</strong></td>
<td><strong>0.586 ± 0.144</strong></td>
<td><strong>0.811 ± 0.232</strong></td>
<td></td>
</tr>
<tr>
<td>multiscale-lagmatrix</td>
<td>0.685 ± 0.332</td>
<td>0.757 ± 0.225</td>
<td>0.555 ± 0.140</td>
<td>0.597 ± 0.168</td>
<td>0.736 ± 0.327</td>
<td></td>
</tr>
<tr>
<td>hierarchical-approx</td>
<td>0.689 ± 0.333</td>
<td>0.757 ± 0.225</td>
<td>0.555 ± 0.140</td>
<td>0.596 ± 0.167</td>
<td>0.736 ± 0.327</td>
<td></td>
</tr>
<tr>
<td>multiscale-haar</td>
<td>0.739 ± 0.318</td>
<td>0.765 ± 0.241</td>
<td>0.533 ± 0.200</td>
<td>0.512 ± 0.200</td>
<td>0.736 ± 0.336</td>
<td></td>
</tr>
</tbody>
</table>
References


Previous work

- Standard offline multiscale anomaly detection using wavelet transform [1], [10], [13], [7].
- Wavelet methods introduce a time delay in the computation of the coefficients at non-dyadic locations which worsens geometrically for coarser scales. Furthermore, they suffer from non-causality, i.e. they need to see some part of the future to assess the presence of an anomaly at present time [6].
- [8] proposed several linear predictive models (Autoregressive, Kalman filter) followed by an anomaly score filtering (by $k\sigma$ rule, or local outlier factor scores introduced by [2]) to detect anomalies.