Detecting Changes in Slope With an L_0 penalty

Sacha HAKIM, Quentin MOAYEDPOUR

M2 MVA - École Normale Supérieure Paris-Saclay
Machine Learning for Time Series

Janvier 2025

Table of contents

1 Context and Motivation

2 Model and Method

3 Application

4 Conclusion

5 References

Context

Change Point Detection (CPD):

- Split a time series into homogeneous sub-series.
- Mathematically: Minimize a segmentation cost.
- Number of change points? Maximum constrain or penalty (or both).

Literature preceding Maidstone, Fearnhead, et al. (2019):

- Binary Segmentation (Scott et al. 1974)
- Segment Neighborhood (Auger et al. 1989)
- Optimal Partitioning (Jackson et al. 2005)
- PELT algorithm (Killick et al. 2012)
- FPOP algorithm (Maidstone, Hocking, et al. 2017)

Motivation

- CPD highly depends on homogeneity metric.
- Fit a piecewise linear function to data? Use a linear cost:

$$\mathcal{C}_{lin}(y_{s+1:t}) = \min_{\beta \in \mathbb{R}^2} \sum_{i=s+1}^{t} (y_i - \beta_0 - \beta_1 i)^2.$$

- What if we also want continuity? (demographic growth, evolution of biomarkers, stock price, ...).
- **Problem:** Continuity imposes the dependence of successive segments in the value of the fitting function. Classical dynamic programming approaches not applicable.
- **Idea of the authors:** Drawing inspiration from PELT and FPOP to build an adapted (and computationally efficient) approach.

Table of contents

1 Context and Motivation

2 Model and Method

3 Application

4 Conclusion

5 References

Model and Cost

- Continuous piecewise linear function fitting problem = parameter estimation problem
- We model $y = \{y_t\}_{t=1}^n$ as the noisy discrete observation of a continuous piecewise linear function:

$$\forall t \in \tau_i + 1, \tau_{i+1},\ y_t \sim \phi_i + \frac{\phi_{i+1} - \phi_i}{\tau_{i+1} - \tau_i}(t - \tau_i) + Z_t,$$

  with Z independent centered noise of (known) variance $\sigma^2 > 0$.
- Estimate the parameters? Define a cost that compromises between good fit and simplicity (needs a penalty).
- We will try to minimize (over τ^m, ϕ):

$$\mathcal{V}(\tau^m, \phi; y) = \sum_{i=0}^{m} \left(\mathcal{C}(y_{\tau_i+1:\tau_i}, \phi_i, \phi_{i+1}) + h(\tau_{i+1} - \tau_i) + \beta\right),$$

  where $\mathcal{C}$ is defined as:

$$\mathcal{C}(y_{s+1:t}, \phi, \psi) = \frac{1}{\sigma^2} \sum_{i=s+1}^{t} \left(y_i - \phi - \frac{\psi - \phi}{t - s}(j - s)\right)^2.$$

Dynamic Programming Approach

- Minimal cost of $y_{1:t}$ for fitting a function with last value ϕ, satisfies a Bellman equation:

$$f^t(\phi) = \min_{0 \leq s < t, \phi' \in \mathbb{R}} f^s(\phi') + \mathcal{C}(y_{s+1:t}, \phi', \phi) + h(t-s) + \beta.$$

- Minimal cost of $y_{1:t}$ for fitting a function with change points τ^k and last value ϕ, also satisfies a Bellman equation:

$$f^t_{\tau^k}(\phi) = \min_{\phi' \in \mathbb{R}} f^{\tau_k}_{\tau_{1:k-1}}(\phi') + \mathcal{C}(y_{\tau_k+1:t}, \phi', \phi) + h(t-\tau_k) + \beta.$$

- Functions $f^t_{\tau^k}$ strictly convex quadratic, computable quadratic coefficients by recurrence.
- **First idea:**
  - Compute optimal values ϕ^*_τ for all change points set τ (recursive minimization of strictly convex quadratics).
  - Minimize $\mathcal{V}(\tau, \phi^*_\tau; y)$ over τ.
  - Too complex...

Pruning Techniques

- Let $\mathcal{T}_t^* = \left\{\tau \in \mathcal{T}_t \,\middle|\, \exists \phi \in \mathbb{R}, f^t(\phi) = f_\tau^t(\phi)\right\}$.
- **Functional pruning:** $\tau \notin \mathcal{T}_s^* \implies \tau \cup \{s\} \notin \mathcal{T}_t^* \quad \forall s < t$.
- **Inequality-based pruning:** Assume $h \geq 0$ and nonincreasing. Set $K = 2\beta + h(1) + h(n)$.

$$\min_{\phi \in \mathbb{R}} f_\tau^s(\phi) > \min_{\phi \in \mathbb{R}} f^s(\phi) + K \implies \min_{\phi \in \mathbb{R}} f_\tau^t(\phi) > \min_{\phi \in \mathbb{R}} f^t(\phi) \quad \forall s < t.$$

- We define $\hat{\mathcal{T}}_t$ recursively:
  - $\hat{\mathcal{T}}_1 \leftarrow \{\emptyset\}$
  - $\hat{\mathcal{T}}_{t+1} \leftarrow \hat{\mathcal{T}}_t \cup \{\tau \cup \{t\} : \tau \in \mathcal{T}_t^*\}$
  - $\hat{\mathcal{T}}_{t+1} \leftarrow \left\{\tau \in \hat{\mathcal{T}}_{t+1} \,\middle|\, \min_{\phi \in \mathbb{R}} f_\tau^t(\phi) \leq \min_{\phi \in \mathbb{R}} f^t(\phi) + K\right\}$

  The two prunings tell us that $\mathcal{T}_t^* \subset \hat{\mathcal{T}}_t$.
- In practice, we can easily compute $\mathcal{T}_t^*$ from $\hat{\mathcal{T}}_t$ with line search. Thus we can compute $\mathcal{T}_n^*$ recursively.

CPOP Algorithm

The authors built the CPOP algorithm from the previous ideas.

- $\hat{\mathcal{T}}_1 \leftarrow \emptyset$.
- Recursively compute $\mathcal{T}_t^*$ with line search and then $\hat{\mathcal{T}}_{t+1}$ from the previous formula.
- The best change points set is within $\mathcal{T}_n^*$. One only needs to minimize $|\mathcal{T}_n^*|$ strictly convex quadratics and compare the minimal values.
- We can then compute the $\phi^*_{\tau^*}$ by minimizing recursively $f^{\tau_k^*}_{\tau^*_{1:k-1}}$.

Table of contents

1 Context and Motivation

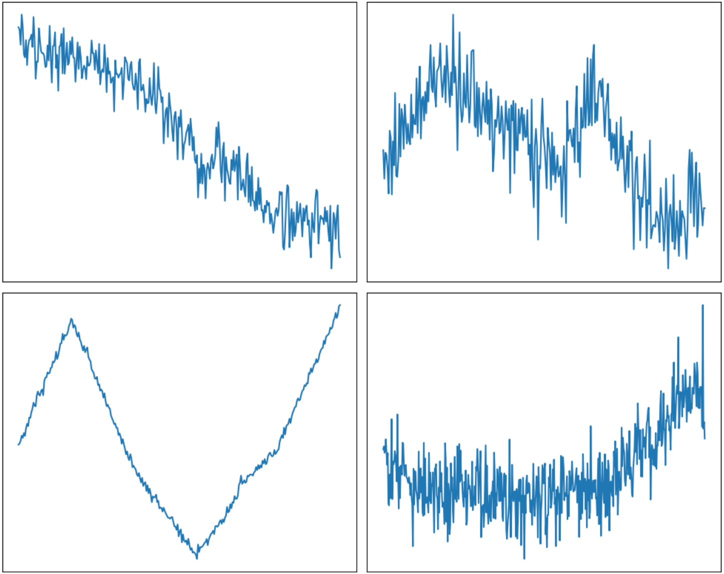
2 Model and Method

3 Application

4 Conclusion

5 References

CPOP on synthetic data



CPOP on synthetic data

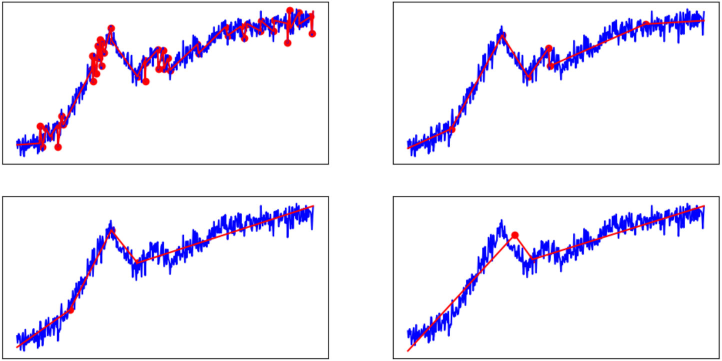
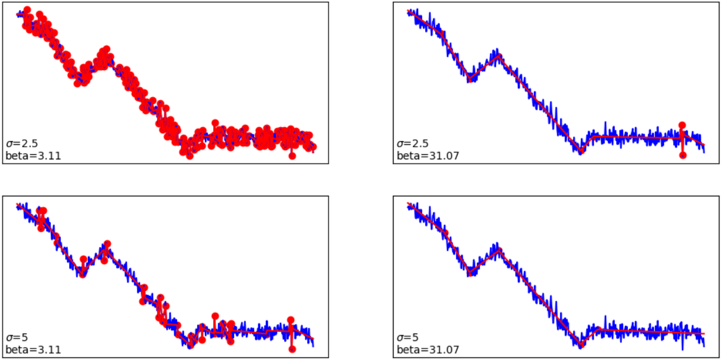


Figure 1: Estimations for different values of σ

CPOP on synthetic data

$$\min \sum_{i=0}^{m} \left[\frac{1}{\sigma^2} \sum_{t=\tau_i+1}^{\tau_{i+1}} \left(y_t - \phi_{\tau_i} - \frac{\phi_{\tau_{i+1}} - \phi_{\tau_i}}{\tau_{i+1} - \tau_i} (t - \tau_i) \right)^2 + h(\tau_{i+1} - \tau_i) \right] + \beta m$$

$$\Leftrightarrow \min \sum_{i=0}^{m} \left[\sum_{t=\tau_i+1}^{\tau_{i+1}} \left(y_t - \phi_{\tau_i} - \frac{\phi_{\tau_{i+1}} - \phi_{\tau_i}}{\tau_{i+1} - \tau_i} (t - \tau_i) \right)^2 + \tilde{h}(\tau_{i+1} - \tau_i) \right] + \tilde{\beta} m$$



Jointly estimate σ and β

Choose a criterion:

- $AIC = -2\mathcal{L}(y|\beta,\sigma) + 2|\tau|$
- $AIC_{L_2} = -2\mathcal{L}(y|\beta,\sigma) + \sigma^2|\tau|$
- $BIC = -2\mathcal{L}(y|\beta,\sigma) + |\tau|log(n)$
- $mBIC = -2\mathcal{L}(y|\beta,\sigma) + 6|\tau|log(n) + \sum_{k=0}^{|\tau|+1} \log\left(\frac{t_{k+1}-t_k}{T}\right)$

Run the algorithm:

Input: $list_\beta$, y
$\sigma_{\text{curr}} \leftarrow \text{MAD}(\text{DIFF}(y))$
for β *in* $list_\beta$ **do**
  detector.run($\sigma = \sigma_{\text{curr}}, \beta$)
  $\sigma_{\text{curr}} \leftarrow$ detector.update_sigma()
  criterion[β] $\leftarrow$ detector.criterion()
end for
$\beta_{\text{min}}, \sigma_{\text{min}} \leftarrow \arg\min_{\beta,\sigma}$ criterion[β]
Return: $(\beta_{\text{min}}, \sigma_{\text{min}})$

Jointly estimate σ and β

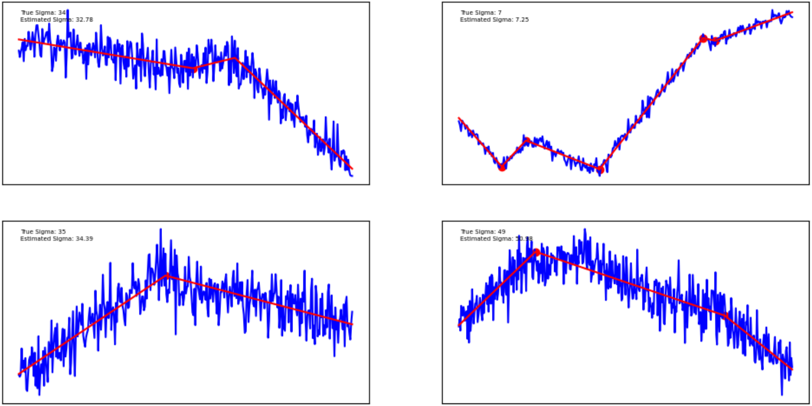
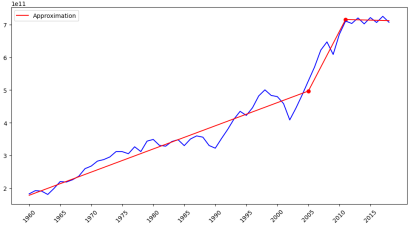
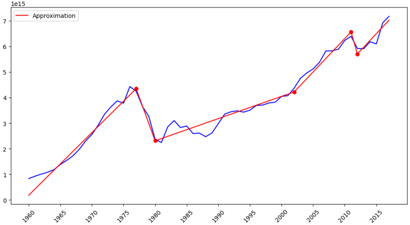


Figure 2: Caption

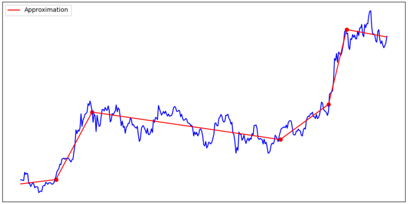
CPOP on real data



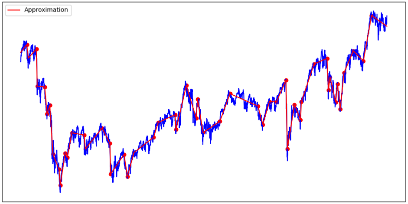
(a) Argentina's GDP



(b) Iran's GDP

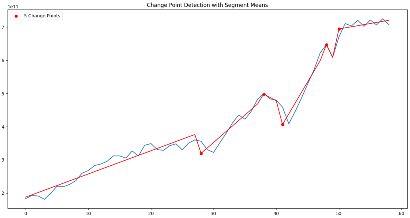


(c) Bitcoin Price.

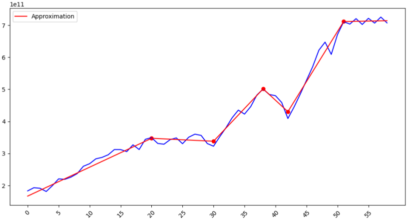


(d) EuroStoxx50 price

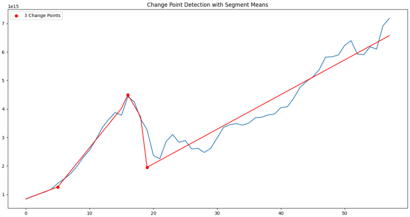
CPOP vs PELT



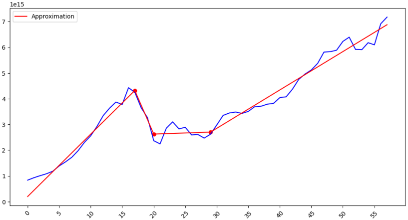
Change Point Detection with Segment Means
1e11
5 Change Points



1e11
Approximation



Change Point Detection with Segment Means
1e15
3 Change Points



1e15
Approximation

Table of contents

1 Context and Motivation

2 Model and Method

3 Application

4 Conclusion

5 References

Conclusion

- The majority of the time was spent on implementing and gaining a clear understanding of the paper
- Clear pseudo-algorithm but limited details are provided on the calculation of the coefficients
- Implementation on Python using dictionnaries to store the coefficients:
  - Not the best for computing the min over a set of keys
  - May suffer from lack of memory on very large time serie
  - In practice, no problems of time or computational ressource
- Statistical criterion for justification of the choice of the parameters
- Not a general model, need to be apply on specific time series (Linear Trends)
- Not robust to outliers

Table of contents

1 Context and Motivation

2 Model and Method

3 Application

4 Conclusion

5 References

References I

- R. Maidstone, P. Fearnhead, et al. (2019). "Detecting Changes in Slope With an L_0 Penalty". In: Journal of Computational and Graphical Statistics 28.2, pp. 265–275.
- R. Maidstone, T. Hocking, et al. (2017). "On Optimal Multiple Changepoint Algorithms for Large Data". In: Statistics and Computing 27.2, pp. 519–533.
- R. Killick et al. (2012). "Optimal Detection of Changepoints with a Linear Computational Cost". In: Journal of the American Statistical Association 107.500, pp. 1590–1598.
- B. Jackson et al. (2005). "An algorithm for optimal partitioning of data on an interval". In: IEEE Signal Processing Letters 12.2, pp. 105–108.
- I. E. Auger et al. (1989). "Algorithms for the Optimal Identification of Segments in Data". In: Bulletin of Mathematical Biology 51.1, pp. 39–54.
- A. J. Scott et al. (1974). "A Cluster Analysis Method for Grouping Means in the Analysis of Variance". In: Biometrics 30.3, pp. 507–512.