APPROXIMATING THE LIKELIHOOD RATIO IN LINEAR-GAUSSIAN STATE-SPACE MODELS FOR CHANGE DETECTION

Kostas Tsampourakis[*] *and Víctor Elvira*[*‡]

[*] School of Mathematics, University of Edinburgh, United Kingdom
[‡] The Alan Turing Institute, United Kingdom

ABSTRACT

Change-point detection methods are widely used in signal processing, primarily for detecting and locating changes in a considered model. An important family of algorithms for this problem relies on the likelihood ratio (LR) test. In state-space models (SSMs), the time series is modeled through a Markovian latent process. In this paper, we focus on the linear-Gaussian (LG) SSM, in which the LR-based methods require running a Kalman filter for every candidate change point. Since the number of candidates grows with the length of the time series, this strategy is inefficient in short time series and unfeasible for long ones. We propose an approximation to the LR which uses a constant number of filters, independently on the time-series length. The approximated LR relies on the Markovian property of the filter, which forgets errors at an exponential rate. We present theoretical results that justify the approximation, and we bound its error. We demonstrate its good performance in two numerical examples.

Index Terms— State-space modeling; change detection.

1. INTRODUCTION

The detection of changes in signals is a problem of broad interest, with applications in biological science [1, 2], climate science [3], speech processing [4], and finance [5], among many others. The problem has had various treatments, stemming from disciplines such as signal processing, industrial process monitoring, control systems, and statistics [6–9]. Change-point detection methods involve assessing how well one or several given statistical models represent an observed signal. They are generally used to detect and locate an abrupt change in the considered model [10] and for further post-processing purposes, e.g., segmentation of signals [11].

State-space models (SSM) provide an efficient framework for modelling stochastic signals with dependent observations and are ubiquitous in science and engineering [12]. In the case of the linear-Gaussian (LG) SSM, exact solutions for the filtering distributions and the marginal likelihood are possible via the Kalman filter [13]. We note that for most SSMs, the marginal likelihood cannot be computed in analytic form and must be estimated via alternative methods such as particle filters [14]. The marginal likelihood plays a central role for statistical tasks such as model estimation, change point detection, outlier detection, etc. [15–19]. A powerful way for detecting a change in the parameters of a model is the *likelihood ratio* (LR) test, that compares between a null hypothesis representing no change and an alternative hypothesis of a single or multiple changes [20]. In this paper, we focus on the case of dependent observations, which has been applied, e.g., in change/fault detection and isolation [21, 22], detection of jumps in linear systems [23], and target tracking [8]. The LR has been extensively studied for independent observations, leading to the CUSUM algorithm [24] (see also [25, 26]).

A practical limitation of the LR procedure in SSMs is that the predictive densities of the observations depend on the location of the change point. As a result, a different filter has to be run for each candidate change point thus resulting in algorithms that require $O(T)$ filters, where T is the number of observations of the time series. This issue has been addressed for the online setting, by limiting the candidate change points to a window of observations [23, 27]. However, this approach cannot be applied to the offline problem, since the change point can occur anywhere in the record. Moreover the performance of the algorithm is sensitive to the window size [27].

In this paper, we present an efficient way of dealing with this problem in LG-SSMs by proposing an *approximate* LR (ALR). The ALR exploits the uniqueness of the steady state convergence of Kalman solutions, and it requires a constant number of filters regardless of the time-series length T. We study this convergence and bound the error of the approximation. Note that, although we consider the case where the post-change model is known, the main ideas can be applied if the post-change parameters are within a set of possible models or even unknown, in which case the *generalized* LR (GLR) test is used [21, 28]. The proposed approximation can be then implemented within GLR-based methods, e.g., *multiple models* algorithm [8], to improve its efficiency.

The rest of the paper is structured as follows. In Section 2, we provide the required background. In Section 3, we introduce the proposed ALR and derive an error bound that guarantees its performance. All proofs are presented in Section 4. The paper concludes with numerical results in Section 5.

The work of V. E. is supported by the ANR under PISCES (ANR-17-CE40-0031-01) and by ARL/ARO under grant W911NF-20-1-0126.

2. BACKGROUND

2.1. Linear State-Space Models and Kalman Filter

Let us consider the following linear dynamical system,

$$\mathbf{x}_t = \mathbf{A}_t \mathbf{x}_{t-1} + \mathbf{q}_t, \tag{1}$$
$$\mathbf{y}_t = \mathbf{H}_t \mathbf{x}_t + \mathbf{r}_t, \tag{2}$$

where $\mathbf{x}_t \in \mathbb{R}^{d_x}$ and $\mathbf{y}_t \in \mathbb{R}^{d_y}$ are the hidden state and observation, respectively, $\mathbf{A}_t$ and $\mathbf{H}_t$ are sequences of known matrices, and $\mathbf{q}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_t)$ and $\mathbf{r}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_t)$ are the state and observation noises, respectively. The system is identified by the parameter $\boldsymbol{\theta}_t = \{\mathbf{A}_t, \mathbf{H}_t, \mathbf{Q}_t, \mathbf{R}_t\}$. When the parameter $\boldsymbol{\theta}_t = \boldsymbol{\theta}, t = 1, \ldots, T$, we refer to it as a static model.

It is possible to compute the predictive and filtering distributions, $p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \mathcal{N}(\mathbf{x}_t|\mathbf{m}_t^-, \mathbf{P}_t^-)$ and $p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \mathcal{N}(\mathbf{x}_t|\mathbf{m}_t, \mathbf{P}_t)$, respectively, with the Kalman filter [13]. We will use the term Kalman step to refer to a single evaluation of the Kalman recursion for a given t. The predictive density of the observation (also called likelihood factor when evaluated at the observation) is $p(\mathbf{y}_t|\mathbf{y}_{1:t-1}) = \mathcal{N}(\mathbf{y}_t|\mathbf{H}_t\mathbf{m}_t^-, \mathbf{S}_t)$.

2.2. Change Detection

The LR is used to decide between a static null hypothesis (no change), $\mathcal{H}^0$, where $\boldsymbol{\theta}_t = \boldsymbol{\theta}^0$ for $t = 1, \ldots, T$, and a set of piecewise static alternative hypotheses (single change), $\mathcal{H}_k$ where $\boldsymbol{\theta}_t = \boldsymbol{\theta}^0$ for $t < k$, and $\boldsymbol{\theta}_t = \boldsymbol{\theta}^1$ for $t \geq k$. Here we assume both pre and post-change parameters are known, but extensions are readily available [20]. The subscripts 0 and 1 will be used for the static models $\mathcal{H}^0$ (with $\boldsymbol{\theta}^0$) and $\mathcal{H}^1$ (with $\boldsymbol{\theta}^1$), respectively. The distributions of interest for the model $\mathcal{H}_k$ will be denoted with superscript (k), e.g., $p^{(k)}(\mathbf{y}_t|\mathbf{y}_{1:t-1})$, denotes the predictive distribution of the observation at time t in a model with change at time k. Note that $p^{(k)}(\mathbf{y}_t|\mathbf{y}_{1:t-1}) = p_0(\mathbf{y}_t|\mathbf{y}_{1:t-1})$ if $k > t$. Then, the joint marginal likelihoods for models $\mathcal{H}^0$, $\mathcal{H}^1$, and $\mathcal{H}_k$ are

$$p_j(\mathbf{y}_{1:T}) = \prod_{t=1}^{T} p_j(\mathbf{y}_t|\mathbf{y}_{1:t-1}), \qquad j \in \{0, 1\}, \tag{3}$$
$$p^{(k)}(\mathbf{y}_{1:T}) = \prod_{t=1}^{k-1} p_0(\mathbf{y}_t|\mathbf{y}_{1:t-1}) \prod_{t=k}^{T} p^{(k)}(\mathbf{y}_t|\mathbf{y}_{1:t-1}), \tag{4}$$

$k \in \{1, ..., T-1\}$, respectively. The evaluation of the joint likelihood in each model requires $O(T)$ Kalman steps.

2.3. Likelihood Ratio

The log-likelihood ratio between $\mathcal{H}_k$ and $\mathcal{H}^0$ is [20]

$$\lambda_{k,T} = \log \frac{p^{(k)}(\mathbf{y}_{1:T})}{p_0(\mathbf{y}_{1:T})} = \sum_{t=k}^{T} \log \frac{p^{(k)}(\mathbf{y}_t|\mathbf{y}_{1:t-1})}{p_0(\mathbf{y}_t|\mathbf{y}_{1:t-1})}. \tag{5}$$

Therefore, it involves computing the marginal likelihoods $p^{(k)}(\mathbf{y}_{1:T})$ for $k = 1, \ldots, T-1$. This requires running one Kalman filter for each value of k, resulting in an algorithm that takes $O(T)$ full filter runs in total, or $O(T^2)$ Kalman steps [8].

3. APPROXIMATE LIKELIHOOD RATIO

3.1. The approximated LR

The computation of the LR in (5) takes a number of Kalman steps that is $O(T^2)$, which poses a computational challenge for long time series. In the proposed approximation, we run only two Kalman filters, one for $\boldsymbol{\theta}^0$ and one for $\boldsymbol{\theta}^1$, achieving a cost of $O(T)$ Kalman steps. The proposed ALR is

$$\tilde{\lambda}_{k,T} = \sum_{t=k}^{T} \log \frac{p_1(\mathbf{y}_t|\mathbf{y}_{1:t-1})}{p_0(\mathbf{y}_t|\mathbf{y}_{1:t-1})}, \tag{6}$$

which is derived by approximating the Kalman solution under model $\mathcal{H}_k$ by the solution under model $\mathcal{H}^1$. Note that, for each k, both models converge to the same solution exponentially fast after the change, as we show in the next section. The approximation errors thus happen mostly in the first few time steps starting from $t = k$.

3.2. Theoretical Properties of the ALR

The approximation error is given by

$$|\lambda_{k,T} - \tilde{\lambda}_{k,T}| = \left| \sum_{t=k}^{T} \log \frac{p^{(k)}(\mathbf{y}_t|\mathbf{y}_{1:t-1})}{p_1(\mathbf{y}_t|\mathbf{y}_{1:t-1})} \right|. \tag{7}$$

In the following, we present two propositions that lead to the main result on the error upper bound, provided in Theorem 1. The first proposition establishes (A) the convergence of the predictive covariance matrices, $\mathbf{P}_t^-$ and $\mathbf{P}_t^{-(k)}$, for $k > t$, to a unique steady-state, and (B) the convergence of the predictive means between the same two models $\mathcal{H}^1$ and $\mathcal{H}_k$. Crucially, (A) and (B) take place at an exponential rate.

Proposition 1. *Under Assumption ($\mathfrak{A}$), there exists a positive semidefinite matrix $\mathbf{P}$, bounded matrix sequences $\tilde{\mathbf{P}}_t$ and $\tilde{\mathbf{P}}_t^{(k)}$, and scalars $M > 0$ and $0 < \beta < 1$, such that*

$$(A) \qquad \mathbf{P}_t^- = \mathbf{P}(\mathbf{I} + \tilde{\mathbf{P}}_t \beta^t), \tag{8}$$
$$\mathbf{P}_t^{-(k)} = \mathbf{P}(\mathbf{I} + \tilde{\mathbf{P}}_t^{(k)} \beta^t), \tag{9}$$
$$(B) \qquad ||\mathbf{m}_t^- - \mathbf{m}_t^{-(k)}||_2 \leq M\beta^{t-k}. \tag{10}$$

Proof: In Section 4, we state Assumption ($\mathfrak{A}$), provide the proof for $d_x = 1$, and discuss the case $d_x > 1$.

The second proposition establishes the convergence of the likelihood factors.

Proposition 2. *Under Assumption ($\mathfrak{A}$), there exist $r > 0$ and $0 < \beta < 1$ such that, for all $t \geq k$, the sequence of error terms satisfies,*

$$\left| \log \frac{p^{(k)}(\mathbf{y}_t|\mathbf{y}_{1:t-1})}{p_1(\mathbf{y}_t|\mathbf{y}_{1:t-1})} \right| \leq r\beta^{t-k}. \tag{11}$$

Proof: In Section 4, we provide proof for $d_x = 1$ and discuss for $d_x > 1$.
This leads to the following bound for the approximation error.

Theorem 1. *An upper bound for the approximation error of the ALR is given by*

$$|\lambda_{k,T} - \tilde{\lambda}_{k,T}| \leq r \sum_{j=1}^{T-k} \beta^{j-1}. \tag{12}$$

Proof: It follows from the definition in (7) *and Proposition 2.*

4. THEORETICAL DERIVATIONS

In this section, we provide explicit proofs of Propositions 1 and 2 for $d_x = 1$, and sketch the proofs for $d_x > 1$.
Proof of Proposition 1. For the case where $d_x = 1$, let us use the lower-case characters a and h for the parameters $\mathbf{A}$ and $\mathbf{H}$ and replace the noise covariance matrices with variances σ_q^2 and σ_r^2. Making the appropriate substitutions in propagation step of the filtering covariance [12, Eq. (4.20)], we obtain for the (scalar) predictive covariance the difference equation

$$P_{t+1}^- = \frac{(a^2\tilde{\sigma}_r^2 + \sigma_q^2)P_t^- + \tilde{\sigma}_r^2\sigma_q^2}{P_t^- + \tilde{\sigma}_r^2}, \tag{13}$$

where $\tilde{\sigma}_r^2 = \sigma_r^2/h^2$. This is the scalar Riccati difference equation with general solution [29]:

$$P_t^- = P(1 + \tilde{P}_t\beta^t), \tag{14}$$

where

$$P = \rho_+ - \tilde{\sigma}_r^2, \tag{15}$$

$$\tilde{P}_t = \frac{\rho_+}{P}\frac{C(\beta - 1)}{1 + C\beta^t}, \tag{16}$$

with $\beta = \frac{\rho_-}{\rho_+}, \rho_\pm = \frac{p}{2} \pm \frac{1}{2}\sqrt{p^2 + 4s}, p = (a^2 + 1)\tilde{\sigma}_r^2 + \sigma_q^2$, and $s = -a^2\tilde{\sigma}_r^4$. The constant C is determined by the initial condition. The part (A) of Proposition 1 is a direct application of (14) for the models $\mathcal{H}^1$ and $\mathcal{H}_k$ and noting that for $t \geq k$ the two models will differ only in the initial condition. As for the part (B), the mean of the predictive distribution of the state for the models $\mathcal{H}^1$ and $\mathcal{H}_k$ can be written as

$$m_t^- = B_{t-1}m_{t-1}^- + aK_{t-1}y_{t-1}, \tag{17}$$

$$m_t^{-(k)} = B_{t-1}^{(k)}m_{t-1}^{-(k)} + aK_{t-1}^{(k)}y_{t-1}, \tag{18}$$

respectively, where $B_t = a(1 - hK_t)$ and K_t is the Kalman gain as in [12, (4.21)]. We unroll the recursions as

$$m_t^- = \prod_{i=1}^{t-k+1} B_{t-i}m_{k-1}^- + \sum_{i=1}^{t-k+1}\left(\prod_{j=1}^{i-1} B_{t-i}\right) aK_{t-i}y_{t-i},$$

$$m_t^{-(k)} = \prod_{i=1}^{t-k+1} B_{t-i}^{(k)}m_{k-1}^{-(k)} + \sum_{i=1}^{t-k+1}\left(\prod_{j=1}^{i-1} B_{t-i}^{(k)}\right) aK_{t-i}^{(k)}y_{t-i}.$$

We introduce now Assumption ($\mathfrak{A}$), which requires that the observation sequence is bounded $|y_t| < \psi$, and that $p + 4s < 0$, and $\rho_+ - \tilde{\sigma}_r^2 > 0$. The latter two conditions depend on the model parameters. In the numerical examples, we have checked that they hold. Then, from the two sequences above,

$$\begin{aligned}|m_t^- - m_t^{-(k)}| \leq & \Big|\prod_{i=1}^{t-k+1} B_{t-i}m_{k-1}^-\Big| + \Big|\prod_{i=1}^{t-k+1} B_{t-i}^{(k)}m_{k-1}^{-(k)}\Big| \\ & + \Big|\sum_{i=1}^{t-k+1}(\prod_{j=1}^{i-1} B_{t-j}^{(k)})aK_{t-i}^{(k)} - (\prod_{j=1}^{i-1} B_{t-j})aK_{t-i}\}y_{t-i}\Big| \\ & \leq T_1 + \sum_{i=1}^{t-k+1} |A_i - \Gamma_i||y_{t-i}|.\end{aligned}$$

where T_1 is the sum of the two leading terms, while A_i and Γ_i are the respective products that appear in the sums. Due to their linear relationship to the Kalman gain and Eq. (14), the sequences B_t and $B_t^{(k)}$ can be written as $B_t = B(1 + \tilde{B}_t\beta^t)$ and $B_t^{(k)} = B(1 + \tilde{B}_t^{(k)}\beta^t)$. Therefore, since both sequences A_i and Γ_i contain products of element, we factor out $|B|^{i-1}$ in each of them. It is possible to see that, in the difference $|A_i - \Gamma_i|$, the constant terms cancel out, while all other terms are proportional to β^{t-i}. Moreover, from Proposition 1(A), it follows that $\tilde{B}_t$ and $\tilde{B}_t^{(k)}$ are bounded, and $\beta < 1$. Hence, there exists a constant M_1 such that $|A_i - \Gamma_i| \leq M_1|B|^{i-1}\beta^{t-i}$. Moreover T_1 equals a bounded sequence times $|B|^{t-k+1}$. Finally, since by Assumption $\mathfrak{A}$, $|B| < \beta$, we obtain

$$\begin{aligned}|m_t^- - m_t^{-(k)}| &\leq M'\beta^{t-k} + \psi M_1 \sum_{i=1}^{t-k+1} |B|^{i-1}\beta^{t-i} \\ &\leq M\beta^{t-k}. \quad \square\end{aligned}$$

Lemma 1. *Let v_t and $u_t > 0$ be sequences in $\mathbb{R}$ whose difference converges to zero as*

$$|v_t - u_t| \leq M\beta^t, \tag{19}$$

for $M > 0$ and $0 < \beta < 1$, and let u_t be bounded, $|u_t| \leq u$. Moreover, consider the positive convergent sequences, $\Omega_t = \Omega(1 + \tilde{\Omega}_t\beta^t), \Omega_t' = \Omega(1 + \tilde{\Omega}_t'\beta^t)$, with $\tilde{\Omega}_t$ and $\tilde{\Omega}_t'$ being bounded and $\Omega > 0$. Then,

$$|\log \Omega_t/\Omega_t'| + |v_t'\Omega_t v_t - u_t'\Omega_t' u_t| = O(\beta^t). \tag{20}$$

Proof. The first term is bounded by using the identity $|\log(1+x)| \leq |x/(1+x)|$, if $x > -1$. For the second term, after

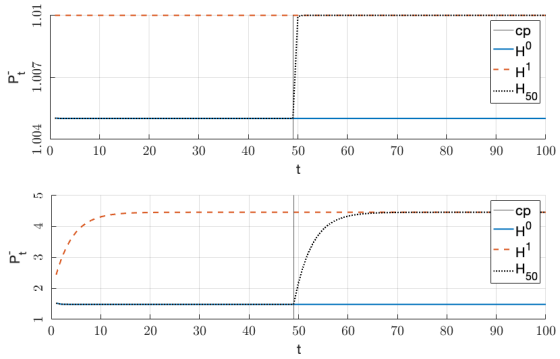

Fig. 1: Evolution of the variance of the predictive filtering distribution, P_t^-, for the true model $\mathcal{H}_{50}$ (black dotted), $\mathcal{H}^0$ (solid blue), and $\mathcal{H}^1$ (dashed red), in Scenario (F) (top) and Scenario (S) (bottom).

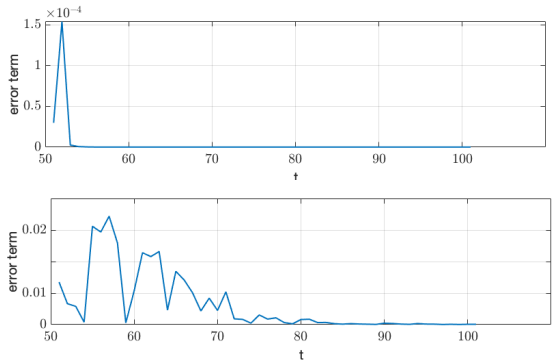

Fig. 2: Error of the log-likelihood between $\mathcal{H}_{50}$ and $\mathcal{H}^1$, as in the LHS of (11), in Scenario (F) (top) and Scenario (S) (bottom).

some manipulation we have

$$|\Omega_t v_t^2 - \Omega_t' u_t^2| \leq \Omega_t |v_t - u_t|^2 + 2\Omega_t |u_t||v_t - u_t| + |\Omega_t - \Omega_t'||u_t|^2 \leq (\Gamma M^2 \beta^t + 2\Gamma u M + |\tilde{\Omega}_t - \tilde{\Omega}_t'|u^2)\Omega\beta^t,$$

where Γ bounds $1 + \tilde{\Omega}_t \beta^t$. Since the parenthesis is bounded in the last inequality, then Eq. (19) holds. □

Proof of Proposition 2. Proposition 2 follows from the application of the preceding Lemma 1 to the sequences m_t^- and $m_t^{-(k)}$ and the sequence of inverse variances $\Omega_t = S_t^{-1}$. The assumptions of the Lemma 1 are satisfied by the conclusions of Proposition 1 and Assumption ($\mathfrak{A}$). □

Sketch of the proof for $dx > 1$. We have presented the proof for $d_x = 1$ due to space limitations. The proof for $d_x > 1$ follows a similar logic. The main variation occurs in the proof of Proposition 1(A), where the scalar Riccati difference equation in (13) must be replaced by its matrix analogue. The solution of the matrix Riccati difference equation is of the form of (8) [30]. The proofs of Propositions 1(B) and 2 are straightforward generalizations of $d_x = 1$, with bounded sequences becoming bounded matrix sequences. We will provide the explicit proof in the extended version of this work. □

5. NUMERICAL EXPERIMENTS

In this section, we show two numerical examples that complement the theoretical results. Since the proofs are made explicit for the case $d_x = 1$ due to space constraints, we focus here on the same setting. It is worth noting that we have observed similar behaviours for a higher-dimensional state space, which will be presented in the extension of this work. We generate the time series with $T = 100$ time steps, inducing a the changepoint at $k = 50$ i.e., the set of model parameters is $\boldsymbol{\theta}^0$ for $1 \leq t < 50$, and $\boldsymbol{\theta}^1$ for $50 \leq t \leq 100$. We consider two scenarios, Scenario (F), which is an example of *fast* convergence and Scenario (S), where the convergence is *slow*. In both scenarios, we set $h = 1$ and $\sigma_q = 1$, and the change occurs in the noise variance, with $\sigma_r^0 = 1$ and $\sigma_r^1 = 100$. We set $a = 0.9$ and $a = 0.1$ for Scenario (S) and (F), respectively.

Figure 1 shows the evolution of the variance, P_t^-, for the true model $\mathcal{H}_{50}$ (black dotted), $\mathcal{H}^0$ (solid blue), and $\mathcal{H}^1$ (dashed red), for both Scenario (F) (top) and Scenario (S) (bottom). We observe that the filter that sets $\boldsymbol{\theta}^1$ from the beginning ($\mathcal{H}^1$) converges to the true variance given by $\mathcal{H}_{50}$, which uses the right model at all time steps (with the change at $k = 50$). The convergence is faster in Scenario (F) than Scenario (S). Figure 2 displays the error of the log-likelihood between $\mathcal{H}_{50}$ and $\mathcal{H}^1$, as in the LHS of (11), for $t \geq k$. These results are in agreement with our theoretical derivations. For instance, in the limit where $\sigma_q/\tilde{\sigma}_r << 1$, the expression for the convergence rate simplifies to $\beta \simeq a^2$. Then, the convergence is quick for a small a, and slow for a close to 1.

6. CONCLUSIONS

In this paper, we have proposed an *approximate likelihood ratio* (ALR) that can be used for change-point detection in LG-SSM. The ALR approximates the standard *likelihod ratio* (LR) by requiring only a constant number of filters instead of $O(T)$. We have characterized the error in terms of an upper bound by studying the convergence of the Kalman solutions under different models. We have provided explicit proofs for the uni-dimensional case, and sketched the proofs for the general case. We have shown in two examples that the proposed ALR retains a good performance, showing that the errors depend on the speed in which the Kalman solutions converge. Future work will provide explicit proofs for the general case and, will evaluate the ALR in a more general setting, for instance in the case where the post-change model is unknown.

References

[1] F. Caron, A. Doucet, and R. Gottardo, "On-line change-point detection and parameter estimation with application to genomic data," *Statistics and Computing*, vol. 22, no. 2, pp. 579–595, 2012.

[2] T. D. Johnson, R. M. Elashoff, and S. J. Harkema, "A bayesian change-point analysis of electromyographic data: Detecting muscle activation patterns and associated applications," *Biostatistics*, vol. 4, no. 1, pp. 143–164, 2003.

[3] J. Reeves, J. Chen, X. L. Wang, R. Lund, and Q. Q. Lu, "A review and comparison of changepoint detection techniques for climate data," *Journal of applied meteorology and climatology*, vol. 46, no. 6, pp. 900–915, 2007.

[4] D. Rybach, C. Gollan, R. Schluter, and H. Ney, "Audio segmentation for speech recognition using segment features," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2009, pp. 4197–4200.

[5] M. Lavielle and G. Teyssiere, "Adaptive detection of multiple change-points in asset price volatility," in *Long memory in economics*, pp. 129–156. Springer, 2007.

[6] P. Fearnhead, "Exact and efficient bayesian inference for multiple changepoint problems," *Statistics and computing*, vol. 16, no. 2, pp. 203–213, 2006.

[7] R. P. Adams and D. J. MacKay, "Bayesian online changepoint detection," *arXiv:0710.3742*, 2007.

[8] F. Gustafsson, *Adaptive filtering and change detection*, vol. 1, Wiley, 2000.

[9] I. Hwang, S. Kim, Y. Kim, and C. E. Seah, "A survey of fault detection, isolation, and reconfiguration methods," *IEEE transactions on control systems technology*, vol. 18, no. 3, pp. 636–653, 2009.

[10] S. Aminikhanghahi and D. J. Cook, "A survey of methods for time series change point detection," *Knowledge and information systems*, vol. 51, no. 2, pp. 339–367, 2017.

[11] C. Truong, L. Oudre, and N. Vayatis, "Selective review of offline change point detection methods," *Signal Processing*, vol. 167, pp. 107299, 2020.

[12] S. Särkkä, *Bayesian filtering and smoothing*, Cambridge University Press, 2013.

[13] R. E. Kalman, "A new approach to linear filtering and prediction problems," 1960.

[14] P. M. Djurić, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, and J. Míguez, "Particle filtering," *IEEE Signal Processing Magazine*, vol. 20, no. 5, pp. 19–38, September 2003.

[15] P. M. Djurić and J. Míguez, "Assessment of nonlinear dynamic models by Kolmogorov–Smirnov statistics," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5069–5079, 2010.

[16] C. Maiz, E. Molanes-Lopez, J. Miguez, and P. Djuric, "A particle filtering scheme for processing time series corrupted by outliers," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4611–4627, 2012.

[17] L. Martino, J. Read, V. Elvira, and F. Louzada, "Cooperative parallel particle filters for online model selection and applications to urban mobility," *Digital Signal Processing*, vol. 60, pp. 172 – 185, 2017.

[18] F. Llorente, L. Martino, D. Delgado, and J. Lopez-Santiago, "Marginal likelihood computation for model selection and hypothesis testing: an extensive review," *arXiv preprint arXiv:2005.08334*, 2020.

[19] E. Chouzenoux and V. Elvira, "Graphem: Em algorithm for blind kalman filtering under graphical sparsity constraints," in *Acoustics, Speech and Signal Processing (ICASSP), 2020 IEEE International Conference on*, 2020, pp. 1–5.

[20] M. Basseville, I. V. Nikiforov, et al., *Detection of abrupt changes: theory and application*, vol. 104, prentice Hall Englewood Cliffs, 1993.

[21] P. Li and V. Kadirkamanathan, "Particle filtering based likelihood ratio approach to fault diagnosis in nonlinear stochastic systems," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 31, no. 3, pp. 337–343, 2001.

[22] C. Andrieu, A. Doucet, S. S. Singh, and V. B. Tadic, "Particle methods for change detection, system identification, and control," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 423–438, 2004.

[23] A. S. Willsky and H. L. Jones, "A generalized likelihood ratio approach to state estimation in linear systems subjects to abrupt changes," in *1974 IEEE Conference on Decision and Control including the 13th Symposium on Adaptive Processes*. IEEE, 1974, pp. 846–853.

[24] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, no. 1/2, pp. 100–115, 1954.

[25] R. Jana and S. Dey, "Change detection in teletraffic models," *IEEE Transactions on Signal Processing*, vol. 48, 2000.

[26] A. Tartakovsky, I. Nikiforov, and M. Basseville, *Sequential analysis: Hypothesis testing and changepoint detection*, CRC Press, 2014.

[27] T. L. Lai, "Sequential changepoint detection in quality control and dynamical systems," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 57, no. 4, pp. 613–644, 1995.

[28] T. S. Lau and W. P. Tay, "Quickest change detection in the presence of a nuisance change," *IEEE Transactions on Signal Processing*, vol. 67, 2019.

[29] V. L. Kocic and G. Ladas, *Global behavior of nonlinear difference equations of higher order with applications*, vol. 256, Springer Science & Business Media, 1993.

[30] G. C. Goodwin and K. S. Sin, *Adaptive filtering prediction and control*, Courier Corporation, 2014.