A Hybrid Lower Bound for Parameter Estimation of Signals With Multiple Change-Points

Lucien Bacharach, Mohammed Nabil El Korso, Alexandre Renaux, Jean-Yves Tourneret

A Hybrid Lower Bound for Parameter Estimation of Signals with Multiple Change-Points

Lucien Bacharach, Mohammed Nabil El Korso, Alexandre Renaux, and Jean-Yves Tourneret

***Abstract*—Change-point estimation has received much attention in the literature as it plays a significant role in several signal processing applications. However, the study of the optimal estimation performance in such context is a difficult task since the unknown parameter vector of interest may contain both continuous and discrete parameters, namely the parameters associated with the noise distribution and the change-point locations. In this paper, we handle this by deriving a lower bound on the mean square error of these continuous and discrete parameters. Specifically, we propose a Hybrid Cramér-Rao–Weiss-Weinstein bound and derive its associated closed-form expressions. Numerical simulations assess the tightness of the proposed bound in the case of Gaussian and Poisson observations.**

I. Introduction

Non-stationary signals are often encountered in signal processing applications. Abrupt changes are a common cause of such non-stationarity. The latter occurs when statistical properties of random observations change abruptly, i.e., quickly with respect to (w.r.t.) the sampling rate. In this scenario, the so-called change-point problem can be divided into two categories: i) on-line processing, in which one aims at detecting or estimating the change-point locations from data received sequentially; ii) off-line processing, in which the (possibly multiple) change-point locations are inferred from a batch of data. Typical applications involving on-line change-point estimation include system monitoring and fault detection [1], whereas off-line processing has been used successfully for astronomical or biomedical and image processing [2]–[4]. This paper focuses on the off-line change-point estimation problem and aims at determining a performance measure for this problem. Note that we are not only interested in the change location estimation, but rather in the *joint* estimation of the change locations *and* the signal parameters on each segment (i.e., between each change).

In the parameter estimation framework, the maximum likelihood estimator (MLE) is a commonly used algorithm due to its interesting asymptotic statistical properties (asymptotic efficiency and normality under mild conditions) [5]. Nevertheless, such regularity conditions are not met in the change-point scenario since the change locations (main parameters of interest) are discrete and thus the corresponding likelihood is not differentiable w.r.t. these parameters. Consequently, this leads to considerable difficulties in characterizing the behavior of the MLE. As an example, the seminal work of Hinkley derived the MLE's asymptotic distribution (semi closed-form) in the case of mean value changes in Gaussian observations [6]. Several decades later, Fotopoulos *et al.* derived the exact asymptotic distribution of the MLE for a single change-point affecting the mean of Gaussian sequences [7]. To the best of our knowledge, this analysis has never been carried out in the case of multiple change-points, which is the main objective of the paper.

A useful alternative to the study of the asymptotic behavior of some estimators is to focus on the second-order moments, by resorting to lower bounds on the mean squared error (MSE). These bounds offer a convenient way to characterize the inherent limitations, in terms of MSE, of parameter estimators for a given model. In this paper, we aim at deriving lower bounds on the MSE of parameter estimators, for signals including multiple change-points. The Cramér-Rao bound (CRB) is the most widely used lower bound in estimation theory since it provides, under some regularity conditions (see [8]), the asymptotic variance of the MLE, and generally leads to interesting closed-form expressions. Because of the discrete nature of the change-points, we remind that the required regularity conditions are not met, and the CRB of these parameters cannot be derived. To overcome this issue, there exist other lower bounds on the MSE which do not require the differentiability of the log-likelihood. One can cite deterministic lower bounds including the family of Barankin bounds [9]–[12], and Bayesian lower bounds such as Weiss-Weinstein bounds [13], [14].

Some of the aforementioned lower bounds have been derived for the change-point estimation problem. Specifically, regarding deterministic lower bounds, Ferrari and Tourneret derived the Chapman-Robbins bound (a specific type of Barankin bound) for a single change-point estimation [15]. This work was later extended by La Rosa *et al.* to the case of multiple change-points [16]. Since these deterministic

L. Bacharach and A. Renaux are with Laboratory of Signals and Systems (L2S), Université Paris-Sud, 91192 Gif-sur Yvette, France.

M. N. El Korso is with Laboratory Energetics, Mechanics and Electromagnetism (LEME), Université Paris Nanterre, 91410 Ville d'Avray, France.

Jean-Yves Tourneret is with the IRIT/INP-ENSEEIHT/TéSA, University of Toulouse, 31071 Toulouse, France.

bounds only give a coarse insight on the change-point estimation behavior, we recently proposed the use of Bayesian lower bounds such as the Weiss-Weinstein bound (WWB), both for a single change [17] and multiple changes [18]. It is worth mentioning that to the best of our knowledge, most of the derived bounds in the literature considered the change-points as the only unknown parameters, meaning that all the other parameters were assumed to be known (e.g., in [18], the means and variances of the Gaussian distributions associated with the different segments were assumed to be known). In this paper, we fill this void. More precisely, we propose a hybrid bound for multiple change-point estimation when the discrete change locations and the continuous parameters of the distributions associated with the different segments are both unknown. To achieve this, we propose the "Hybrid Cramér-Rao–Weiss-Weinstein bound" (HCRWWB, later abbreviated as HB for "hybrid bound", for sake of simplicity), in which the CRB part is associated with the continuous parameters, and the WWB part is associated with the discrete parameters. Note that a Bayesian Cramér-Rao–Weiss-Weinstein bound was proposed in [19] in a fully Bayesian context. In contrast to this bound, whose determination required a numerical integration, we investigate in this paper an HB, which can be determined using closed-form expressions.

This paper is structured as follows: in Section II, we expose the observation model used throughout the paper. In Section III, we give the general expression of the proposed bound. This bound is derived for some specific change-point problems in Section IV, namely, for Gaussian and Poisson observations. Simulation results illustrating the interest of the proposed bound are presented in Section V. Finally, our conclusions and future works are reported in Section VI.

II. Problem Statement

This section introduces the observation model used throughout this paper. We consider a time series $\boldsymbol{x} = [x_1, \ldots, x_N] \in \Omega$ with independent random variables $x_n \in \Omega'$ ($\Omega \subset \mathbb{R}^N$ denotes the observation space for $\boldsymbol{x}$, and Ω' denotes the observation space for one observation x_n, i.e., $\Omega = (\Omega')^N = \Omega' \times \ldots \times \Omega'$). This time series $\boldsymbol{x}$ is subjected to multiple abrupt changes, that arise at unknown time instants $t_1, \ldots, t_Q$, also referred to as change-points or changes. The total number of changes Q is assumed to be known. Thus, the observation model can be written as

$$\begin{cases} x_n \sim f(x_n\,; \boldsymbol{\eta}_0) & \text{for } n = 1, \ldots, t_1, \\ \quad\vdots & \qquad\quad\vdots \\ x_n \sim f(x_n\,; \boldsymbol{\eta}_Q) & \text{for } n = t_Q + 1, \ldots, N, \end{cases} \tag{1}$$

in which $f(x_n\,; \boldsymbol{\eta}_q)$ denotes the distribution of the random variables x_n in the $(q+1)$-th segment, namely the segment delineated by the two consecutive change-points t_q and t_{q+1}, with $q \in [\![0\,, Q]\!]$ (that is the set of integers between 0 and Q), $t_0 \triangleq 0$ and $t_{Q+1} \triangleq N$. These distributions $f(.\,; \boldsymbol{\eta}_q)$ are parameterized by parameter vectors $\boldsymbol{\eta}_q = [\eta_{q_1}, \ldots, \eta_{q_L}]^\mathsf{T} \in \mathbb{R}^L$ (for instance, in the Gaussian case, $L = 2$ and the parameter vector $\boldsymbol{\eta}_q$ includes the mean and the variance of the Gaussian distribution for the $(q+1)$-th segment). We assume that all the distributions $f(.\,; \eta_q)$, for $q \in [\![0\,, Q]\!]$, belong to the same family.

The parameter estimation problem in such scenario consists in estimating i) the change-point locations t_q, $q = 1, \ldots, Q$, and ii) the parameter vectors $\boldsymbol{\eta}_q$, $q = 0, \ldots, Q$. Thus, the unknown parameter vector to estimate is $\boldsymbol{\theta} = [\boldsymbol{\eta}^\mathsf{T}, \boldsymbol{t}^\mathsf{T}]^\mathsf{T} \in \Theta \subset \mathbb{R}^{L(Q+1)+Q}$, with $\boldsymbol{\eta} = [\boldsymbol{\eta}_0^\mathsf{T}, \ldots, \boldsymbol{\eta}_Q^\mathsf{T}]^\mathsf{T} \in \Theta_{\boldsymbol{\eta}} \subset \mathbb{R}^{L(Q+1)}$ and $\boldsymbol{t} = [t_1, \ldots, t_Q]^\mathsf{T} \in \Theta_{\boldsymbol{t}} \subset \mathbb{Z}^Q$. The sets Θ, $\Theta_{\boldsymbol{\eta}}$ and $\Theta_{\boldsymbol{t}}$ denote the parameter spaces for $\boldsymbol{\theta}$, $\boldsymbol{\eta}$ and $\boldsymbol{t}$, respectively, that is, $\Theta = \Theta_{\boldsymbol{\eta}} \times \Theta_{\boldsymbol{t}}$.

The purpose of this paper is to assess estimation performance for the parameter vector $\boldsymbol{\theta}$ by providing lower bounds on the mean squared error (MSE) for a family of estimators $\hat{\boldsymbol{\theta}}(\boldsymbol{x})$ of $\boldsymbol{\theta}$. As explained in Section I, we provided in [18] a lower bound on the estimation error for the vector $\boldsymbol{t}$ only (parameter vectors $\boldsymbol{\eta}_q$ were assumed to be known), assuming $\boldsymbol{t}$ is a random vector (Bayesian lower bound). In [19], addressing a specific type of data (namely, a time-series with Poisson entries), the (scalar) parameters η_q were assumed unknown and random, i.e., a fully Bayesian point of view for the estimation of $\boldsymbol{t}$ and $\boldsymbol{\eta}$ was considered. In this paper, we fill the gap between [18] and [19] by generalizing the work presented in [19] to any pre-specified signal distribution $f(.\,; \boldsymbol{\eta}_q)$ (parameterized by unknown vectors $\boldsymbol{\eta}_q$; see Section IV-A). In this sense, the bound derived in this paper is more general – for applications to specific usual distributions, see Section IV-B. Another difference is the estimation framework: while in [19] we used a fully Bayesian point of view, we now consider a hybrid context, in the sense that the parameter vectors $\boldsymbol{\eta}_q$ stacked in $\boldsymbol{\eta}$ are assumed unknown and deterministic, with true values $\boldsymbol{\eta}_q^\star$ (accordingly, the true value of the full parameter vector is denoted by $\boldsymbol{\eta}^\star$), and the parameter vector $\boldsymbol{t}$ is assumed random. Consequently, the estimator $\hat{\boldsymbol{\theta}}(\boldsymbol{x})$ is hybrid as well, for example it can be the ML-MAP estimator (Maximum Likelihood-Maximum *A Posteriori*) [20, p. 12], [21]. The context of hybrid estimation is appropriate in cases where no *a priori* information is available on some of the unknown parameters. In addition, interestingly, the hybrid point of view makes it possible to get rid of some integrals that were evaluated numerically in [19]. Finally, the interest of the hybrid set-up lies in the resulting trade-off between tightness of the bound and its computational complexity: thanks to the contribution of the Weiss-Weinstein bound, which is known to be one of the tightest bounds, the proposed bound is tight, but it also entails a reasonable computational cost, as closed-form expressions for its elements can be obtained (see Section IV).

Since the parameter vector $\boldsymbol{t}$ is random, in agreement with the Bayesian framework, it is assigned a prior distribution denoted by $\pi(\boldsymbol{t})$. Note that this distribution is assumed to be independent of $\boldsymbol{\eta}$. In this paper, since the number of changes Q is assumed to be known, we choose a convenient prior that is compatible with this assumption: we assume that the change-points are drawn according to a uniform random walk. In other words, t_q is assumed to be given by $t_q = t_{q-1} + \epsilon_q$, $q = 1, \ldots, Q$, where $t_0 \triangleq 0$ and ϵ_q are i.i.d. uniformly distributed random variables on the set of integers $[\![d\,, D]\!]$. The values of d and D can be freely chosen as long as $d \geq 1$, $d < D$, and the last change t_Q occurs at least before the end of the observation window, i.e., the maximum possible value for D is $D_{\max} = \lfloor (N-1)/Q \rfloor$, with $\lfloor . \rfloor$ denoting the floor function. Thus, the joint (discrete) prior distribution for the parameter vector $\boldsymbol{t}$ is given by

$$\pi(\boldsymbol{t}) = \frac{1}{\Delta^Q} \prod_{q=1}^{Q} \mathbb{1}_{[\![t_{q-1}+d, t_{q-1}+D]\!]}(t_q), \tag{2}$$

in which we have defined $\Delta \triangleq D - d + 1$. The support of this prior distribution is denoted by $\mathcal{T}' \triangleq \left\{\boldsymbol{t} = [t_1, \ldots, t_Q]^\mathsf{T} \in \mathbb{Z}^Q \,\middle|\, \forall q \in [\![1\,, Q]\!], t_q \in [\![t_{q-1} + d\,, t_{q-1} + D]\!],\ t_0 = 0,\ t_Q < N\right\}$, and it corresponds to the set of the possible segmentations of the observation window $[\![1\,, N]\!]$ into exactly Q segments, with maximum length D and minimum length d. This prior distribution offers an interesting trade-off between the exploration of the parameter space and the computational complexity of the resulting bound. The results of this paper could be extended to other prior distributions, without any guarantee that closed-form expressions of the bound would be obtained.

From model (1) and the aforementioned assumptions, the likelihood of the observations can be written as

$$f(\boldsymbol{x}\,|\,\boldsymbol{t}\,;\boldsymbol{\eta}) = \prod_{q=0}^{Q} \prod_{n=t_q+1}^{t_{q+1}} f(x_n\,;\boldsymbol{\eta}_q), \tag{3}$$

with $t_0 \triangleq 0$ and $t_{Q+1} \triangleq N$. Note that from (2) and (3), it is also possible to write the joint distribution between the observations $\boldsymbol{x}$ and the parameter vector $\boldsymbol{t}$, for some given parameter vector $\boldsymbol{\eta}$, as $f(\boldsymbol{x}, \boldsymbol{t}\,;\boldsymbol{\eta}) \triangleq f(\boldsymbol{x}\,|\,\boldsymbol{t}\,;\boldsymbol{\eta})\,\pi(\boldsymbol{t})$.

As already mentioned, the number of changes Q is assumed to be known and the estimation of Q is beyond the scope of this paper. The estimation of Q is often referred to as a model dimension estimation problem (see e.g., [22], [23]). In array processing, the model dimension corresponds to the number of sources [24], and it is very classical to assume that it is known [25]. In our problem, since Q determines the size of the unknown parameter vector, it implies a strong link between Q, $t_1, \ldots, t_Q$ and $\boldsymbol{\eta}_0, \ldots, \boldsymbol{\eta}_Q$, and it becomes necessary to fix the value of Q in order to assess the estimation performance of $\boldsymbol{\theta}$. As we will see hereafter, although Q is known, the derivation of lower bounds provides important information on how difficult the estimation problem is.

Note that the random variable x_n can be either absolutely continuous or discrete, depending on the application. In the following, we will assume that it is continuous. However, the extension to the discrete case is straightforward: an example of discrete observations is investigated in Section IV-B2 of this paper. We now present the lower bound on the MSE which we derive thereafter.

III. Proposed Bound

This section presents the lower bound on the mean square error derived in this paper (the derivation itself, for the problem introduced in Section II, is carried out in Section IV). We first recall a general inequality leading to the proposed lower bound, namely the covariance inequality.

A. *Background on the covariance inequality*

We consider an estimation problem, with an unknown parameter vector $\boldsymbol{\theta} \in \Theta$ which can be either deterministic, or random, or hybrid – the latter case being considered in this paper. Let $\hat{\boldsymbol{\theta}}(\boldsymbol{x})$ an estimator of $\boldsymbol{\theta}$, i.e., a measurable function $\Omega \to \Theta$. Let $\boldsymbol{v}(\boldsymbol{x}, \boldsymbol{\theta})$ a real measurable function, such that (i) the covariance matrix $\mathbb{E}\{\boldsymbol{v}(\boldsymbol{x}, \boldsymbol{\theta})\boldsymbol{v}^\mathsf{T}(\boldsymbol{x}, \boldsymbol{\theta})\}$ is positive definite, and (ii) the matrices $\mathbb{E}\{\hat{\boldsymbol{\theta}}(\boldsymbol{x})\boldsymbol{v}^\mathsf{T}(\boldsymbol{x}, \boldsymbol{\theta})\}$ and $\mathbb{E}\{\boldsymbol{\theta}\boldsymbol{v}^\mathsf{T}(\boldsymbol{x}, \boldsymbol{\theta})\}$ have finite elements. The following matrix inequality then holds and is commonly referred to as the covariance inequality [5], [20]:

$$\mathbb{E}\left\{\left(\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta}\right)^\mathsf{T}\right\} \succcurlyeq \boldsymbol{C}\boldsymbol{V}^{-1}\boldsymbol{C}^\mathsf{T}, \tag{4}$$

in which $\boldsymbol{C} \triangleq \mathbb{E}\{(\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta})\boldsymbol{v}^\mathsf{T}(\boldsymbol{x}, \boldsymbol{\theta})\}$, $\boldsymbol{V} \triangleq \mathbb{E}\{\boldsymbol{v}(\boldsymbol{x}, \boldsymbol{\theta})\boldsymbol{v}^\mathsf{T}(\boldsymbol{x}, \boldsymbol{\theta})\}$ and the matrix inequality $\boldsymbol{A} \succcurlyeq \boldsymbol{B}$ denotes the so-called Löwner partial ordering, i.e., the difference $\boldsymbol{A} - \boldsymbol{B}$ is a nonnegative matrix. As explained in details in [20], the covariance inequality (4) corresponds to the vector extension of the Cauchy-Schwarz inequality (see also [20, p. 33] for a proof of (4)).

At this point, it is worth noticing that, without any further assumptions on the vector function $\boldsymbol{v}(\boldsymbol{x}, \boldsymbol{\theta})$, the matrix $\boldsymbol{C}$ in (4) generally depends on $\hat{\boldsymbol{\theta}}(\boldsymbol{x})$, hence the right-hand side of (4) is *not* an interesting lower bound on the MSE of $\hat{\boldsymbol{\theta}}(\boldsymbol{x})$. However, for some well-chosen vector functions $\boldsymbol{v}(\boldsymbol{x}, \boldsymbol{\theta})$, and for some adequate set of estimators $\hat{\boldsymbol{\theta}}$, it is possible to obtain a matrix $\boldsymbol{C}$ that does not depend on $\hat{\boldsymbol{\theta}}(\boldsymbol{x})$. In that case, the right-hand side of (4) is a lower bound on the MSE and applies to *any* estimator in the aforementioned set. We now explain how we define the vector functions $\boldsymbol{v}(\boldsymbol{x}, \boldsymbol{\theta})$ to obtain the lower bound derived in this paper. For other examples explaining how to define $\boldsymbol{v}(\boldsymbol{x},\boldsymbol{\theta})$ and lower bounds stemming from the covariance inequality, see [20, pp. 35–53].

B. The hybrid Cramér-Rao–Weiss-Weinstein bound (HCR-WWB)

In order to make the formulation of the bound more generic, we use notations that slightly differ from those in Section II. Let us consider an R-dimensional hybrid unknown parameter vector $\boldsymbol{\theta} = [\boldsymbol{\theta}_{\mathsf{d}}^{\mathsf{T}}, \boldsymbol{\theta}_{\mathsf{r}}^{\mathsf{T}}]^{\mathsf{T}}$ belonging to the parameter space $\Theta \subset \mathbb{R}^R$. Note that the term "hybrid" here means that $\boldsymbol{\theta}_{\mathsf{d}}$ is deterministic, belonging to a subset Π_{d} of $\mathbb{R}^P$, and whose true value is $\boldsymbol{\theta}_{\mathsf{d}}^{\star}$, while $\boldsymbol{\theta}_{\mathsf{r}}$ is random, belonging to $\mathbb{R}^Q$ (such that $R = P+Q$). For a given value of the deterministic parameter vector $\boldsymbol{\theta}_{\mathsf{d}}$, we define a prior distribution $\pi(\boldsymbol{\theta}_{\mathsf{r}}\,;\boldsymbol{\theta}_{\mathsf{d}})$ for the random parameter vector $\boldsymbol{\theta}_{\mathsf{r}}$, whose support is denoted by $\Pi_{\mathsf{r}} \subset \mathbb{R}^Q$. Note that the prior might explicitly depend on $\boldsymbol{\theta}_{\mathsf{d}}$ in some cases, see [26]. Let us denote by $f(\boldsymbol{x}\,|\,\boldsymbol{\theta}_{\mathsf{r}}\,;\boldsymbol{\theta}_{\mathsf{d}})$ the likelihood of the observations, so that the function $f(\boldsymbol{x},\boldsymbol{\theta}) \triangleq f(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}}\,;\boldsymbol{\theta}_{\mathsf{d}}) = f(\boldsymbol{x}\,|\,\boldsymbol{\theta}_{\mathsf{r}}\,;\boldsymbol{\theta}_{\mathsf{d}})\,\pi(\boldsymbol{\theta}_{\mathsf{r}}\,;\boldsymbol{\theta}_{\mathsf{d}})$ denotes the joint probability density function (p.d.f.) of $\boldsymbol{x}$ and $\boldsymbol{\theta}_{\mathsf{r}}$ parameterized by $\boldsymbol{\theta}_{\mathsf{d}}$. The following relations can be established between these notations and those from Section II: $\boldsymbol{\theta}_{\mathsf{d}} \equiv \boldsymbol{\eta}$, $\boldsymbol{\theta}_{\mathsf{r}} \equiv \boldsymbol{t}$, $P \equiv L(Q+1)$, $Q \equiv Q$, $R \equiv L(Q+1)+Q$, $\pi(\boldsymbol{\theta}_{\mathsf{r}}\,;\boldsymbol{\theta}_{\mathsf{d}}) = \pi(\boldsymbol{\theta}_{\mathsf{r}}) \equiv \pi(\boldsymbol{t})$ and $\Pi_{\mathsf{r}} \equiv \mathscr{T}'$. Let us define $\Theta' \triangleq \{\boldsymbol{\theta} \in \Theta \mid f(\boldsymbol{x},\boldsymbol{\theta}) > 0,$ for almost all $\boldsymbol{x} \in \Omega\}$. The statistical expectation of a vectorial functional $\boldsymbol{g}_{\boldsymbol{\theta}_{\mathsf{d}}}(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}})$ parameterized by $\boldsymbol{\theta}_{\mathsf{d}}$, w.r.t. the joint p.d.f. $f(\boldsymbol{x},\boldsymbol{\theta}) = f(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}}\,;\boldsymbol{\theta}_{\mathsf{d}})$ is denoted by $\mathbb{E}_{\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}};\boldsymbol{\theta}_{\mathsf{d}}}\{\boldsymbol{g}_{\boldsymbol{\theta}_{\mathsf{d}}}(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}})\}$. Finally, let us denote by $\hat{\boldsymbol{\theta}} \triangleq \hat{\boldsymbol{\theta}}(\boldsymbol{x}) = [\hat{\boldsymbol{\theta}}_{\mathsf{d}}^{\mathsf{T}}, \hat{\boldsymbol{\theta}}_{\mathsf{r}}^{\mathsf{T}}]^{\mathsf{T}}$ a joint (hybrid) estimator of $[\boldsymbol{\theta}_{\mathsf{d}}^{\mathsf{T}}, \boldsymbol{\theta}_{\mathsf{r}}^{\mathsf{T}}]^{\mathsf{T}}$, i.e., $\hat{\boldsymbol{\theta}}_{\mathsf{d}} \triangleq \hat{\boldsymbol{\theta}}_{\mathsf{d}}(\boldsymbol{x})$ is an estimator of the deterministic parameter vector $\boldsymbol{\theta}_{\mathsf{d}}^{\star}$ and $\hat{\boldsymbol{\theta}}_{\mathsf{r}} \triangleq \hat{\boldsymbol{\theta}}_{\mathsf{r}}(\boldsymbol{x})$ is an estimator of a realization of the random vector $\boldsymbol{\theta}_{\mathsf{r}}$.

In order to obtain a lower bound on the estimation error of the hybrid parameter vector $\boldsymbol{\theta}$, the idea is to combine two different lower bounds w.r.t. $\boldsymbol{\theta}_{\mathsf{d}}$ and $\boldsymbol{\theta}_{\mathsf{r}}$ respectively. This results in a "hybrid" lower bound for the estimation of the parameter vector $\boldsymbol{\theta}$. Such kind of lower bound has already been proposed in the literature [27]–[29]. In this paper, we propose to combine the (deterministic) Cramér-Rao bound [30], [31, Chap. 32], [8, p. 300] with the (Bayesian) Weiss-Weinstein bound [13]. A fully Bayesian version of this bound was first proposed in a recursive form in [32], and was recently adapted to the off-line change-point problem for Poisson data in [19].

In order to derive the proposed hybrid bound, the vector function $\boldsymbol{v}(\boldsymbol{x},\boldsymbol{\theta}) : \Omega \times \Theta \to \mathbb{R}^R$ is constructed in two parts since the estimation is hybrid: the P first components of $\boldsymbol{v}(\boldsymbol{x},\boldsymbol{\theta})$ are related to the deterministic parameters, while the Q last components of $\boldsymbol{v}(\boldsymbol{x},\boldsymbol{\theta})$ $(P+Q=R)$ are related to the random parameters. For $p = 1,\ldots,P$, we set [33]

$$[\boldsymbol{v}(\boldsymbol{x},\boldsymbol{\theta})]_p = \begin{cases} \left.\frac{\partial \ln f(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}};\boldsymbol{\theta}_{\mathsf{d}})}{\partial \theta_{\mathsf{d},p}}\right|_{\boldsymbol{\theta}_{\mathsf{d}}=\boldsymbol{\theta}_{\mathsf{d}}^{\star}}, & \text{if } \boldsymbol{\theta} \in \Theta' \\ 0, & \text{if } \boldsymbol{\theta} \notin \Theta', \end{cases} \tag{5}$$

in which $\theta_{\mathsf{d},p}$ denotes the p-th component of the deterministic parameter vector $\boldsymbol{\theta}_{\mathsf{d}}$, and the derivatives are evaluated at the true value $\boldsymbol{\theta}_{\mathsf{d}}^{\star}$ of the parameter vector $\boldsymbol{\theta}_{\mathsf{d}}$. For $q = P+1,\ldots,R$, we set [13], [29]

$$[\boldsymbol{v}(\boldsymbol{x},\boldsymbol{\theta})]_q = \begin{cases} \frac{f^{s_q}(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}}+\boldsymbol{h}_q;\boldsymbol{\theta}_{\mathsf{d}}^{\star})}{f^{s_q}(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}};\boldsymbol{\theta}_{\mathsf{d}}^{\star})} - \frac{f^{1-s_q}(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}}-\boldsymbol{h}_q;\boldsymbol{\theta}_{\mathsf{d}}^{\star})}{f^{1-s_q}(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}};\boldsymbol{\theta}_{\mathsf{d}}^{\star})}, & \text{if } \boldsymbol{\theta} \in \Theta' \\ 0, & \text{if } \boldsymbol{\theta} \notin \Theta', \end{cases} \tag{6}$$

in which $s_q \in \left]0\,,1\right[$, and the vector $\boldsymbol{h}_q$ is constrained to belong to the set $\mathscr{H}_{\boldsymbol{\theta}_{\mathsf{r}}} \triangleq \left\{\boldsymbol{h} \in \mathbb{R}^Q \,\middle|\, \boldsymbol{\theta}_{\mathsf{r}} + \boldsymbol{h} \in \Pi_{\mathsf{r}}\right\}$. Any values of $s_q \in \left]0\,,1\right[$ and $\boldsymbol{h}_q \in \mathscr{H}_{\boldsymbol{\theta}_{\mathsf{r}}}$ lead to a lower bound for the MSE, but not necessarily the tightest (see Section III-C for a method to find suitable values of s_q and $\boldsymbol{h}_q$).

Without any further assumption, the matrix $\boldsymbol{C}$ in (4) still depends on $\hat{\boldsymbol{\theta}}(\boldsymbol{x})$, which is unwanted. We then make the two following additional assumptions:

1- We consider the class of estimators that are unbiased w.r.t. $\boldsymbol{\theta}_{\mathsf{d}}$, i.e.,

$$\mathbb{E}_{\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}};\boldsymbol{\theta}_{\mathsf{d}}^{\star}}\{\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta}|_{\boldsymbol{\theta}_{\mathsf{d}}^{\star}}\} = [\boldsymbol{0}^{\mathsf{T}}, \boldsymbol{d}^{\mathsf{T}}]^{\mathsf{T}} \tag{7}$$

in which $\boldsymbol{d}$ is an arbitrary vector with size Q, independent of $\boldsymbol{\theta}$.

2- As for the classical CRB, we assume that, for any $\boldsymbol{\theta}_{\mathsf{d}} \in \Pi_{\mathsf{d}}$,

$$\mathbb{E}_{\boldsymbol{x}|\boldsymbol{\theta}_{\mathsf{r}};\boldsymbol{\theta}_{\mathsf{d}}}\left\{\frac{\partial \ln f(\boldsymbol{x}\,|\,\boldsymbol{\theta}_{\mathsf{r}};\boldsymbol{\theta}_{\mathsf{d}})}{\partial \boldsymbol{\theta}_{\mathsf{d}}}\right\} = \boldsymbol{0}. \tag{8}$$

Thus, assuming that both conditions (7) and (8) are satisfied, it can be proved that the matrix $\boldsymbol{C}$ is block diagonal and given by $\boldsymbol{C} = \mathrm{bdiag}\{\boldsymbol{I}, \boldsymbol{C}_{22}\}$, where $\boldsymbol{I}$ denotes the $P \times P$ identity matrix and the columns of $\boldsymbol{C}_{22}$ are given, for $1 \leqslant q \leqslant Q$ and $\boldsymbol{\theta} \in \Theta'$, by

$$\begin{aligned} \boldsymbol{c}_q &= \mathbb{E}_{\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}};\boldsymbol{\theta}_{\mathsf{d}}^{\star}}\Bigg\{\left[\hat{\boldsymbol{\theta}}_{\mathsf{r}}(\boldsymbol{x}) - \boldsymbol{\theta}_{\mathsf{r}}\right] \\ &\quad \times \left[\frac{f^{s_q}(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}}+\boldsymbol{h}_q;\boldsymbol{\theta}_{\mathsf{d}}^{\star})}{f^{s_q}(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}};\boldsymbol{\theta}_{\mathsf{d}}^{\star})} - \frac{f^{1-s_q}(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}}-\boldsymbol{h}_q;\boldsymbol{\theta}_{\mathsf{d}}^{\star})}{f^{1-s_q}(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}};\boldsymbol{\theta}_{\mathsf{d}}^{\star})}\right]\Bigg\} \\ &= \boldsymbol{h}_q\,\mu(s_q,\boldsymbol{h}_q), \end{aligned} \tag{9}$$

in which $\hat{\boldsymbol{\theta}}_{\mathsf{r}}(\boldsymbol{x})$ denotes the estimate of $\boldsymbol{\theta}_{\mathsf{r}}$ obtained by selecting the appropriate components of $\hat{\boldsymbol{\theta}}(\boldsymbol{x})$. In (9), we also have defined, for $\boldsymbol{h}_{\mathsf{r}} \in \mathscr{H}_{\boldsymbol{\theta}_{\mathsf{r}}}$ and $s \in \left]0\,,1\right[$,

$$\begin{aligned} \mu(s,\boldsymbol{h}_{\mathsf{r}}) &\triangleq \iint\limits_{\Omega\times\mathbb{R}^Q} f^s(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}}+\boldsymbol{h}_{\mathsf{r}}\,;\boldsymbol{\theta}_{\mathsf{d}}^{\star})\,f^{1-s}(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}}\,;\boldsymbol{\theta}_{\mathsf{d}}^{\star})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{\theta}_{\mathsf{r}} \\ &= \mathbb{E}_{\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}};\boldsymbol{\theta}_{\mathsf{d}}^{\star}}\left\{\frac{f^s(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}}+\boldsymbol{h}_{\mathsf{r}};\boldsymbol{\theta}_{\mathsf{d}}^{\star})}{f^s(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}};\boldsymbol{\theta}_{\mathsf{d}}^{\star})}\right\}. \end{aligned} \tag{10}$$

The proof that the upper-left block in $\boldsymbol{C}$ is the identity can be found in [8, Sec. 4.3.3.1], while (9) is obtained by using the change of variables $\boldsymbol{\theta}'_{\mathsf{r}} = \boldsymbol{\theta}_{\mathsf{r}} - \boldsymbol{h}_q$. The same type of considerations leads to the expression of the matrix $\boldsymbol{V}$:

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{V}_{11} & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{12}^{\mathsf{T}} & \boldsymbol{V}_{22} \end{bmatrix}, \tag{11}$$

in which

i) (block $\boldsymbol{V}_{11}$) for $(p,p') \in [\![1\,,P]\!]^2$ and $\boldsymbol{\theta} \in \Theta'$, the element $[\boldsymbol{V}_{11}]_{p,p'}$ in the matrix $\boldsymbol{V}$ is given by

$$\begin{aligned}[\boldsymbol{V}_{11}]_{p,p'} &= \mathbb{E}_{\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}};\boldsymbol{\theta}_{\mathsf{d}}^{\star}}\left\{ \frac{\partial \ln f(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}};\boldsymbol{\theta}_{\mathsf{d}})}{\partial \theta_{\mathsf{d},p}}\bigg|_{\boldsymbol{\theta}_{\mathsf{d}}^{\star}} \frac{\partial \ln f(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}};\boldsymbol{\theta}_{\mathsf{d}})}{\partial \theta_{\mathsf{d},p'}}\bigg|_{\boldsymbol{\theta}_{\mathsf{d}}^{\star}} \right\} \\ &= -\,\mathbb{E}_{\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}};\boldsymbol{\theta}_{\mathsf{d}}^{\star}}\left\{ \frac{\partial^2 \ln f(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}};\boldsymbol{\theta}_{\mathsf{d}})}{\partial \theta_{\mathsf{d},p}\partial \theta_{\mathsf{d},p'}}\bigg|_{\boldsymbol{\theta}_{\mathsf{d}}^{\star}} \right\},\end{aligned} \tag{12}$$

in which the last equality holds only if $f(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}};\boldsymbol{\theta}_{\mathsf{d}})$ is twice differentiable w.r.t. the vector $\boldsymbol{\theta}$. Note that this block corresponds to the so-called "modified Cramér-Rao lower bound" for the estimation error of an unbiased estimator of $\boldsymbol{\theta}_{\mathsf{d}}$ [34];

ii) (block $\boldsymbol{V}_{22}$) for $(q,q') \in [\![P+1\,,R]\!]^2$ and $\boldsymbol{\theta} \in \Theta'$,

$$\begin{aligned}[\boldsymbol{V}_{22}]_{q,q'} &= \mathbb{E}_{\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}};\boldsymbol{\theta}_{\mathsf{d}}^{\star}}\left\{ \left[\frac{f^{s_q}(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}}+\boldsymbol{h}_q;\boldsymbol{\theta}_{\mathsf{d}}^{\star})}{f^{s_q}(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}};\boldsymbol{\theta}_{\mathsf{d}}^{\star})} - \frac{f^{1-s_q}(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}}-\boldsymbol{h}_q;\boldsymbol{\theta}_{\mathsf{d}}^{\star})}{f^{1-s_q}(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}};\boldsymbol{\theta}_{\mathsf{d}}^{\star})} \right] \right. \\ &\qquad \left. \times \left[\frac{f^{s_{q'}}(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}}+\boldsymbol{h}_{q'};\boldsymbol{\theta}_{\mathsf{d}}^{\star})}{f^{s_{q'}}(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}};\boldsymbol{\theta}_{\mathsf{d}}^{\star})} - \frac{f^{1-s_{q'}}(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}}-\boldsymbol{h}_{q'};\boldsymbol{\theta}_{\mathsf{d}}^{\star})}{f^{1-s_{q'}}(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}};\boldsymbol{\theta}_{\mathsf{d}}^{\star})} \right] \right\} \\ &= \xi(s_q, s_{q'}, \boldsymbol{h}_q, \boldsymbol{h}_{q'}) + \xi(1-s_q, 1-s_{q'}, -\boldsymbol{h}_q, -\boldsymbol{h}_{q'}) \\ &\quad - \xi(s_q, 1-s_{q'}, \boldsymbol{h}_q, -\boldsymbol{h}_{q'}) - \xi(1-s_q, s_{q'}, -\boldsymbol{h}_q, \boldsymbol{h}_{q'}),\end{aligned} \tag{13}$$

in which we have defined

$$\begin{aligned}&\xi(\alpha,\beta,\boldsymbol{h}_a,\boldsymbol{h}_b) \\ &\triangleq \mathbb{E}_{\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}};\boldsymbol{\theta}_{\mathsf{d}}^{\star}}\left\{ \frac{f^{\alpha}(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}}+\boldsymbol{h}_a;\boldsymbol{\theta}_{\mathsf{d}}^{\star})\, f^{\beta}(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}}+\boldsymbol{h}_b;\boldsymbol{\theta}_{\mathsf{d}}^{\star})}{f^{\alpha+\beta}(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}};\boldsymbol{\theta}_{\mathsf{d}}^{\star})} \right\}.\end{aligned} \tag{14}$$

This block corresponds to the usual Weiss-Weinstein lower bound for the estimation error associated with an estimator of $\boldsymbol{\theta}_{\mathsf{r}}$ [13]. Note that $\mu(s,\boldsymbol{h}) = \xi(s,0,\boldsymbol{h},\boldsymbol{0})$. Thus, only the calculation of $\xi(\alpha,\beta,\boldsymbol{h}_a,\boldsymbol{h}_b)$ is required for the determination of the elements in this block.

iii) (block $\boldsymbol{V}_{12}$) for $(p,q) \in [\![1\,,P]\!] \times [\![P+1\,,R]\!]$ and $\boldsymbol{\theta} \in \Theta'$

$$\begin{aligned}[\boldsymbol{V}_{12}]_{p,q} &= \mathbb{E}_{\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}};\boldsymbol{\theta}_{\mathsf{d}}^{\star}}\left\{ \frac{\partial \ln f(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}};\boldsymbol{\theta}_{\mathsf{d}})}{\partial \theta_{\mathsf{r},p}}\bigg|_{\boldsymbol{\theta}_{\mathsf{d}}^{\star}} \right. \\ &\quad \left. \times \left[\frac{f^{s_q}(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}}+\boldsymbol{h}_q;\boldsymbol{\theta}_{\mathsf{d}}^{\star})}{f^{s_q}(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}};\boldsymbol{\theta}_{\mathsf{d}}^{\star})} - \frac{f^{1-s_q}(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}}-\boldsymbol{h}_q;\boldsymbol{\theta}_{\mathsf{d}}^{\star})}{f^{1-s_q}(\boldsymbol{x},\boldsymbol{\theta}_{\mathsf{r}};\boldsymbol{\theta}_{\mathsf{d}}^{\star})} \right] \right\}.\end{aligned} \tag{15}$$

This block corresponds to the cross-terms between the Cramér-Rao and the Weiss-Weinstein lower bounds.

Finally, for each value of $\boldsymbol{H} \triangleq [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_Q]$ and $\boldsymbol{s} \triangleq [s_1, \ldots, s_Q]^{\mathsf{T}}$, we obtain a lower bound $\boldsymbol{W}(\boldsymbol{H},\boldsymbol{s}) \triangleq \boldsymbol{C}\boldsymbol{V}^{-1}\boldsymbol{C}^{\mathsf{T}}$ on the mean square error. The proposed bound, that is the hybrid Cramér-Rao–Weiss-Weinstein bound (HCR-WWB, from now on abbreviated as HB), is defined as the tightest of these lower bounds;

$$\mathbf{HB} = \sup_{\substack{\boldsymbol{H} \in \mathscr{H}_{\boldsymbol{\theta}_{\mathsf{r}}}^{Q} \\ \boldsymbol{s} \in]0,1[^{Q}}} \boldsymbol{W}(\boldsymbol{H},\boldsymbol{s}), \tag{16}$$

where the order relation underlying the supremum operation is the Löwner ordering. Since the Löwner ordering is only a partial ordering, the uniqueness of the supremum is not guaranteed. Consequently, the maximization operation required in (16) is not trivial and can be very time consuming. For this reason, we describe in the next section a method to make this computation feasible, leading to a suitable HB.

C. Practical computation of the bound

Without any further assumption, obtaining closed-form expressions for $\boldsymbol{W}(\boldsymbol{H},\boldsymbol{s})$ in (16) and computing the supremum are infeasible tasks, even for simple problems. In order to overcome this issue, two solutions are commonly adopted: (i) the set $\mathscr{H}_{\boldsymbol{\theta}_{\mathsf{r}}}^{Q}$ is restricted to diagonal matrices $\boldsymbol{H}$, i.e., the components of $\boldsymbol{h}_q$ are all zero, except its q-th element [13], [16]; and (ii) it has been noticed after extensive numerical experiments that the value $s_j = 0.5$, for all $j \in [\![1\,,Q]\!]$, leads to the tightest WWB [13], [20, p. 41], [35], which reduces the task dimensionality by a factor Q. We therefore suggest to set this value for the HB as well. It is also worth noticing that, in the change-point context, the set $\mathscr{H}_{\boldsymbol{\theta}_{\mathsf{r}}}^{Q}$ is countable and finite, since it is defined from the (discrete) set $\mathcal{T}'$ ($\Pi_{\mathsf{r}} \equiv \mathcal{T}'$ for the change-point problem).

For some given value of $\boldsymbol{s} \in]0\,,1[^{Q}$ (e.g., $s_q = 0.5$, $\forall q$, as suggested), the supremum operation is computed, w.r.t. the Löwner partial ordering, over the set of matrices $\mathscr{W}_{\boldsymbol{s}} \triangleq \{\boldsymbol{W}(\boldsymbol{H},\boldsymbol{s}) \in \mathbb{S}_{+}^{R}\,;\, \boldsymbol{H} \in \mathscr{H}_{\boldsymbol{\theta}_{\mathsf{r}}}^{Q}\}$, in which $\mathbb{S}_{+}^{R}$ denotes the set of nonnegative matrices with size $R \times R$. Note that $\mathscr{W}_{\boldsymbol{s}}$ is a discrete and finite subset of $\mathbb{S}_{+}^{R}$. As stated above, the uniqueness of the supremum of $\mathscr{W}_{\boldsymbol{s}}$ might not be guaranteed. However, if a unique supremum exists, it may:

– either belong to $\mathscr{W}_{\boldsymbol{s}}$, in which case it is the greatest element;

– or not belong to $\mathscr{W}_{\boldsymbol{s}}$, in which case it is the least element in the set of upper bounds to $\mathscr{W}_{\boldsymbol{s}}$.

Otherwise, a unique supremum does not exist, but the set $\mathscr{W}_{\boldsymbol{s}}$ may have several maximal elements ($\boldsymbol{A}$ is a maximal element of $\mathscr{W}_{\boldsymbol{s}}$ if there is no $\boldsymbol{A}' \in \mathscr{W}_{\boldsymbol{s}}$ such that $\boldsymbol{A}' \succ \boldsymbol{A}$). In such a case, we have to find a minimal element in the set of upper bounds of $\mathscr{W}_{\boldsymbol{s}}$. Formally, such a minimal upper bound $\boldsymbol{B}$ verifies $\boldsymbol{B} \succcurlyeq \mathscr{W}_{\boldsymbol{s}}$ (upper bound) and, if there exists a smaller element $\boldsymbol{B}'$ such that $\boldsymbol{B} \succcurlyeq \boldsymbol{B}' \succcurlyeq \mathscr{W}_{\boldsymbol{s}}$, then necessarily $\boldsymbol{B}' = \boldsymbol{B}$ (minimal). Here again, $\boldsymbol{B}$ may not be unique without any additional constraint (if it is unique, then it is the unique least upper bound of $\mathscr{W}_{\boldsymbol{s}}$, that is its unique supremum).

Despite the previous comments, following [16, Sect. III.D.], one way of obtaining a suitable minimal upper bound of $\mathscr{W}_{\boldsymbol{s}}$, defined in a unique manner, is explained hereafter. For each matrix $\boldsymbol{A} \in \mathbb{S}_+^R$, one can associate a centered hyper-ellipsoid, defined as the set $\mathscr{E}(\boldsymbol{A}) \triangleq \{\boldsymbol{y} \in \mathbb{R}^R \,|\, \boldsymbol{y}^\mathsf{T}\boldsymbol{A}^{-1}\boldsymbol{y} \leqslant 1\}$. Referring to [16, Lemma 3], one can show that, for any positive definite matrices $\boldsymbol{A}$ and $\boldsymbol{A}'$, we have the equivalence: $\boldsymbol{A} \preccurlyeq \boldsymbol{A}'$ iff $\mathscr{E}(\boldsymbol{A}) \subseteq \mathscr{E}(\boldsymbol{A}')$. Thus, given any finite family of matrices $\{\boldsymbol{A}_i\}_{i\in I}$ (with I some finite set), one can find the minimum volume hyper-ellipsoid, denoted $\mathscr{E}(\boldsymbol{A}_{\mathrm{LJ}})$, that covers the union of the hyper-ellipsoids $\mathscr{E}(\boldsymbol{A}_i)$, associated with matrices $\boldsymbol{A}_i$. The minimum volume ellipsoid $\mathscr{E}(\boldsymbol{A}_{\mathrm{LJ}})$ is called the Löwner-John ellipsoid for $\bigcup_{i\in I}\mathscr{E}(\boldsymbol{A}_i)$, and one can show that the matrix $\boldsymbol{A}_{\mathrm{LJ}}$ is a minimal upper bound of the set $\{\boldsymbol{A}_i \,;\, i \in I\}$ (see [16, Theorem 4]). Finding $\boldsymbol{A}_{\mathrm{LJ}}$ corresponds to a convex optimization problem [36, p. 411] that can be solved efficiently using a semidefinite programming procedure, such as those provided by the CVX package [37].

Hence, to approach the HB defined in (16), we compute a suitable minimal upper bound of $\mathscr{W}_{\boldsymbol{s}}$, namely the matrix $\boldsymbol{W}_{\mathrm{LJ}}$ associated with the Löwner-John ellipsoid covering the set $\bigcup_{\boldsymbol{H}\in\mathscr{H}^Q_{\boldsymbol{\theta}_r}}\mathscr{E}(\boldsymbol{W}(\boldsymbol{H},\boldsymbol{s}))$.

We have presented the general expression of the bound, that is applicable to any hybrid estimation problem, and have explained how this bound can be computed. We now give its expression for the problem exposed in Section II.

IV. Expressions of the HB for the change-point problem

A. General case

Let us derive the expression of the HB for signals modeled by (1), i.e., signals which include a given number Q of abrupt changes. We first do it in the general case, i.e., for any distribution of the observations $f(.;\boldsymbol{\eta}_q)$ (assumed to be known). We proceed in the following way: first, we derive the block $\boldsymbol{V}_{11}$ as given by (12). Then, we compute the block $\boldsymbol{V}_{12}$ in (13). Finally, we give the general expression of $\xi(\alpha,\beta,\boldsymbol{h}_a,\boldsymbol{h}_b)$ from (14), which enables us to obtain the expressions of blocks $\boldsymbol{C}_{22}$ and $\boldsymbol{V}_{22}$ according to equations (9) and (13), respectively.

1) Block $\boldsymbol{V}_{11}$: In this section, we give the expression of the elements of block $\boldsymbol{V}_{11}$ for signals which include abrupt changes, i.e., those associated with model (1). Rewriting equation (12) with notations from Section II, and setting the index changes $p = Lq + \ell$ and $p' = Lq' + \ell'$, with $(q,q') \in [\![0\,,Q]\!]^2$ and $(\ell,\ell') \in [\![1\,,L]\!]^2$, we have $[\boldsymbol{V}_{11}]_{p,p'} = -\,\mathbb{E}_{\boldsymbol{x},\boldsymbol{t};\boldsymbol{\eta}^\star}\left\{\left.\frac{\partial^2 \ln f(\boldsymbol{x},\boldsymbol{t};\boldsymbol{\eta})}{\partial\eta_{q,\ell}\partial\eta_{q',\ell'}}\right|_{\boldsymbol{\eta}=\boldsymbol{\eta}^\star}\right\}$. In Appendix A, we show that the matrix block $\boldsymbol{V}_{11}$ is block-diagonal, i.e., can be written

$$\boldsymbol{V}_{11} = \mathrm{diag}\Bigg(\frac{d+D}{2}\,\boldsymbol{F}(\boldsymbol{\eta}_0^\star),\ldots,\frac{d+D}{2}\,\boldsymbol{F}(\boldsymbol{\eta}_{Q-1}^\star),\\ \left(N-\frac{Q(d+D)}{2}\right)\boldsymbol{F}(\boldsymbol{\eta}_Q^\star)\Bigg), \quad (17)$$

in which $\boldsymbol{F}(\boldsymbol{\eta}_q)$ is the $(L\times L)$ Fisher information matrix for $\boldsymbol{\eta}_q \in \mathbb{R}^L$, and for one observation in the q-th segment, i.e., for $q \in [\![0\,,Q]\!]$ and $(\ell,\ell') \in [\![1\,,L]\!]^2$,

$$\left[\boldsymbol{F}(\boldsymbol{\eta}_q)\right]_{\ell,\ell'} = -\,\mathbb{E}_{x;\boldsymbol{\eta}_q}\left\{\frac{\partial^2 \ln f(x\,;\boldsymbol{\eta})}{\partial\eta_{q,\ell}\partial\eta_{q,\ell'}}\right\} \\ = -\int_{\Omega'}\frac{\partial^2 \ln f(x\,;\boldsymbol{\eta}_q)}{\partial\eta_{q,\ell}\partial\eta_{q,\ell'}}f(x\,;\boldsymbol{\eta}_q)\,\mathrm{d}x. \quad (18)$$

Invoking the block diagonal structure of the block matrix $\boldsymbol{V}_{11}$ leads to a decoupling between parameter vectors $\boldsymbol{\eta}_q$ associated with each segment.

2) Blocks $\boldsymbol{V}_{22}$ and $\boldsymbol{C}_{22}$: As mentioned earlier, this block corresponds to the Weiss-Weinstein bound on the MSE of the parameter vector $\boldsymbol{t}$ for given parameters $\boldsymbol{\eta}_0,\ldots,\boldsymbol{\eta}_Q$. In other words, we follow exactly the same methodology as in [18]. The main result of that paper states that the block $\boldsymbol{V}_{22}$ is tridiagonal, i.e., for any $(q,q') \in [\![1\,,Q]\!]^2$ such that $|q-q'|>1$, we have

$$[\boldsymbol{V}_{22}]_{q,q'} = 0. \quad (19)$$

The diagonal terms of $\boldsymbol{V}_{22}$ correspond to the numerator of (26) in [18], i.e., are defined as

$$[\boldsymbol{V}_{22}]_{q,q} = u_{\mathsf{D}}(\Delta,h_q) \\ \times\left(\rho_q^{|h_q|}(\varepsilon_{h_q}(2s_q)) + \rho_q^{|h_q|}(\varepsilon_{h_q}(2s_q-1))\right) \\ -\,2\,u_{\mathsf{D}}(\Delta,2h_q)\cdot\rho_q^{2|h_q|}(\varepsilon_{h_q}(s_q)), \quad (20)$$

in which, i) for $\Delta \in \mathbb{N}$, $q \in [\![1\,,Q]\!]$ and $h_q \in \mathbb{Z}$,

$$u_{\mathsf{D}}(\Delta,h_q) \triangleq \begin{cases} \dfrac{(\Delta-|h_q|)^2}{\Delta^2} & \text{if } q<Q \text{ and } |h_q|<\Delta \\ \dfrac{\Delta-|h_Q|}{\Delta} & \text{if } q=Q \text{ and } |h_Q|<\Delta \\ 0 & \text{if } |h_q|\geqslant\Delta, \end{cases} \quad (21)$$

ii) for $q \in [\![1\,,Q]\!]$ and $s \in \,]0\,,1[$,

$$\rho_q(s) \triangleq \int_{\Omega'} f^s(x\,;\boldsymbol{\eta}_{q-1}^\star)\,f^{1-s}(x\,;\boldsymbol{\eta}_q^\star)\,\mathrm{d}x, \quad (22)$$

and iii) for $h \in \mathbb{Z}$ and $s \in \,]0\,,1[$,

$$\varepsilon_h(s) \triangleq \begin{cases} s & \text{if } h>0 \\ 1-s & \text{if } h<0. \end{cases} \quad (23)$$

From the terms in the block $\boldsymbol{V}_{22}$, we can compute the block $\boldsymbol{C}_{22}$ since we have seen that $\mu(s,\boldsymbol{h}) = \xi(s,0,\boldsymbol{h},\boldsymbol{0})$. In addition, due to the structure chosen for vectors $\boldsymbol{h}_q$, namely only its q-th component is nonzero, the block matrix $\boldsymbol{C}_{22}$ is diagonal. The resulting expression of $\mu(s, \boldsymbol{h})$ corresponds to equation (15) in [18]. After plugging this expression into (9), we obtain

$$[\boldsymbol{C}_{22}]_{q,q} = h_q\, u_{\mathsf{D}}(\Delta, h_q)\, \rho_q^{|h_q|}(\varepsilon_{h_q}(s_q)). \quad (24)$$

Finally, the remaining nonzero terms in the block $\boldsymbol{V}_{22}$ are the superdiagonal ones (which equal, by symmetry, the subdiagonal ones). Referring to [18, equations (31) and (32)], the superdiagonal terms in the block $\boldsymbol{V}_{22}$ are given, for $q \in [\![1\,, Q-1]\!]$, by

$$\begin{aligned}[\boldsymbol{V}_{22}]_{q,q+1} = {}& \operatorname{sign}(h_q\, h_{q+1})\, u_{\mathsf{S}}(\Delta, h_q, h_{q+1}) \\ & \times \Upsilon_q(\boldsymbol{d}, s_q, s_{q+1}, h_q, h_{q+1}) \\ & \times \rho_q^{|h_q|}(\varepsilon_{h_q}(s_q))\, \rho_{q+1}^{|h_{q+1}|}(\varepsilon_{h_{q+1}}(s_{q+1})),\end{aligned} \quad (25)$$

in which we have used the following definitions: i) $\boldsymbol{d} \triangleq [d, D]^{\mathsf{T}}$, ii) for $q \in [\![1\,, Q-1]\!]$ and $(h_q, h_{q+1}) \in \mathbb{Z}^2$,

$$u_{\mathsf{S}}(\Delta, h_q, h_{q+1}) \quad (26)$$
$$\triangleq \begin{cases} \dfrac{(\Delta - |h_q|)(\Delta - |h_{q+1}|)}{\Delta^3} & \text{if } q < Q \\ & \text{and } \max(|h_q|, |h_{q+1}|) < \Delta \\ \dfrac{\Delta - |h_Q|}{\Delta^2} & \text{if } q = Q \text{ and } |h_Q| < \Delta \\ 0 & \text{if } \max(|h_q|, |h_{q+1}|) \geqslant \Delta, \end{cases}$$

iii) for $q \in [\![1\,, Q-1]\!]$ and $(s, s') \in \,]0\,, 1[^2$,

$$r_q(s, s') \triangleq \int_{\Omega'} f^{s}(x\,; \boldsymbol{\eta}^\star_{q-1})\, f^{s'}(x\,; \boldsymbol{\eta}^\star_{q})\, f^{1-s-s'}(x\,; \boldsymbol{\eta}^\star_{q+1})\, \mathrm{d}x, \quad (27)$$

iv) for $q \in [\![1\,, Q-1]\!]$ and $(s, s') \in \,]0\,, 1[^2$,

$$R_q(s, s') \triangleq \frac{\rho_q(s)\, \rho_{q+1}(s')}{r_q(s, s'-s)}, \quad (28)$$

and v) for $q \in [\![1\,, Q-1]\!]$, $(h_q, h_{q+1}) \in \mathbb{Z}^2$ and $(s_q, s_{q+1}) \in \,]0\,, 1[^2$, and defining $(x)^+ \triangleq \max(x, 0)$ $x \in \mathbb{R}$,

$$\begin{aligned}\Upsilon_q(\boldsymbol{d}, s_q, s_{q+1}, h_q, h_{q+1}) \triangleq {}& 2(\Delta - |h_q| - |h_{q+1}|)^+ \\ & - (D - |h_q| - |h_{q+1}| + 1)^+ - (\Delta - \max(|h_q|, |h_{q+1}|))^+ \\ & - \frac{1 - R_q^{d - \min(|h_q|, |h_{q+1}|)}\big(\varepsilon_{h_q}(s_q), \varepsilon_{h_{q+1}}(s_{q+1})\big)}{1 - R_q\big(\varepsilon_{h_q}(s_q), \varepsilon_{h_{q+1}}(s_{q+1})\big)},\end{aligned} \quad (29)$$

if $\min(|h_q|, |h_{q+1}|) \geqslant d+1$, or

$$\begin{aligned}\Upsilon_q(\boldsymbol{d}, s_q, s_{q+1}, h_q, h_{q+1}) \triangleq {}& 2(\Delta - |h_q| - |h_{q+1}|)^+ \\ & - 2(\Delta - \max(|h_q|, |h_{q+1}|))^+,\end{aligned} \quad (30)$$

if $\min(|h_q|, |h_{q+1}|) \leqslant d$.

3) Block $\boldsymbol{V}_{12}$: In this section, we are interested in the elements of the block $\boldsymbol{V}_{12}$. For $p \in [\![1\,, L(Q+1)]\!]$ and $q \in [\![1\,, Q]\!]$, and setting $p = L\tilde{q} + \ell$ with $\tilde{q} \in [\![0\,, Q]\!]$ and $\ell \in [\![1\,, L]\!]$, (15) can be rewritten as

$$\begin{aligned}[\boldsymbol{V}_{12}]_{p,q} = {}& \mathbb{E}_{\boldsymbol{x},\boldsymbol{t};\boldsymbol{\eta}^\star}\Bigg\{ \frac{\partial \ln f(\boldsymbol{x}, \boldsymbol{t}; \boldsymbol{\eta})}{\partial \eta_{\tilde{q},\ell}}\Bigg|_{\boldsymbol{\eta}^\star} \\ & \times \left[\frac{f^{s_q}(\boldsymbol{x}, \boldsymbol{t} + \boldsymbol{h}_q\,; \boldsymbol{\eta}^\star)}{f^{s_q}(\boldsymbol{x}, \boldsymbol{t}\,; \boldsymbol{\eta}^\star)} - \frac{f^{1-s_q}(\boldsymbol{x}, \boldsymbol{t} - \boldsymbol{h}_q\,; \boldsymbol{\eta}^\star)}{f^{1-s_q}(\boldsymbol{x}, \boldsymbol{t}\,; \boldsymbol{\eta}^\star)} \right] \Bigg\}.\end{aligned} \quad (31)$$

As shown in Appendix B, the matrix block $\boldsymbol{V}_{12}$ has the form

$$\boldsymbol{V}_{12} = \begin{bmatrix} \boldsymbol{v}_1 & \dots & \boldsymbol{0} \\ \boldsymbol{w}_1 & \ddots & \vdots \\ \vdots & \ddots & \boldsymbol{v}_Q \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{w}_Q \end{bmatrix}, \quad (32)$$

where, for $q \in [\![1\,, Q]\!]$, the $L \times 1$ vectors $\boldsymbol{v}_q$ and $\boldsymbol{w}_q$ have components that can be written, for $\ell \in [\![1\,, L]\!]$, on the one hand, as (see (B.58))

$$\begin{aligned}v_{q,\ell} = [\boldsymbol{V}_{12}]_{L(q-1)+\ell,q} = {}& -h_q\, u_{\mathsf{D}}(\Delta, h_q) \\ & \times \rho_q^{|h_q|-1}(\varepsilon_{h_q}(s_q))\, \varphi_{\eta_{q-1,\ell},q}(\varepsilon_{h_q}(s_q)),\end{aligned} \quad (33)$$

and, on the other hand, (see (B.59))

$$\begin{aligned}w_{q,\ell} = [\boldsymbol{V}_{12}]_{Lq+\ell,q} = {}& h_q\, u_{\mathsf{D}}(\Delta, h_q) \\ & \times \rho_q^{|h_q|-1}(\varepsilon_{h_q}(s_q))\, \varphi_{\eta_{q,\ell},q}(\varepsilon_{h_q}(s_q)),\end{aligned} \quad (34)$$

where, for $j \in [\![1\,, Q]\!]$, $\tilde{j} \in \{j-1, j\}$, $k \in [\![1\,, L]\!]$, and $s \in \,]0\,, 1[$, $\varphi_{\eta_{\tilde{j},k},j}(s)$ is defined by

$$\varphi_{\eta_{\tilde{j},k},j}(s) \triangleq \int_{\Omega'} \frac{\partial \ln f(x\,; \boldsymbol{\eta}_{\tilde{j}})}{\partial \eta_{\tilde{j},k}}\Bigg|_{\boldsymbol{\eta}^\star} f^{s}(x\,; \boldsymbol{\eta}^\star_{j-1})\, f^{1-s}(x\,; \boldsymbol{\eta}^\star_{j})\, \mathrm{d}x. \quad (35)$$

To conclude, equations (17), (19), (20), (24), (25), (32), (33) and (34) provide all the expressions necessary to determine the elements of the matrix $\boldsymbol{W}(\boldsymbol{H}, \boldsymbol{s}) \triangleq \boldsymbol{C}\boldsymbol{V}^{-1}\boldsymbol{C}^{\mathsf{T}}$ in (16). It is worth noticing that, due to the structure of the matrices $\boldsymbol{V}_{11}$, $\boldsymbol{V}_{12}$ and $\boldsymbol{V}_{22}$, the inversion of $\boldsymbol{V}$ should not be particularly difficult from a computational point of view.

In the next section, we give more explicit expressions of these elements for widely encountered distributions in signal processing applications, namely, Gaussian [1], [3], [38]–[40] and Poisson [2], [41]–[43] distributions.

B. Gaussian and Poisson distributions

For each of these cases, we give closed-form expressions for i) $\boldsymbol{F}(\boldsymbol{\eta}_q)$, defined in (18) and which leads to (17); ii) $\rho_q(s)$, defined in (22), which directly leads to (20) and (24), and partly to (25), (33) and (34); iii) $R_q(s, s')$, defined in (27), which leads to (25); and finally for iv) $\varphi_{\eta_{q,\ell},q'}(s)$, defined in (35), which leads to (33) and (34).

1) Gaussian case: The Gaussian distribution has perhaps the widest range of applications (see for instance [1], [3], [7], and references therein for an overview of the potential applications). In such cases, the model (1) is relevant: we can consider that the signal is piecewise Gaussian, i.i.d., that is, for $q \in [\![0\,,Q]\!]$ and $n \in [\![t_q+1\,,t_{q+1}]\!]$ (with $t_0 \triangleq 0$ and $t_{Q+1} \triangleq N$), we have $x_n \sim \mathcal{N}(\mu_q, \sigma_q^2)$. In other words, both mean and variance are likely to change from one segment to another, which means that $L = 2$ and the parameter vector $\boldsymbol{\eta}_q \triangleq [\mu_q, \sigma_q^2]^{\mathsf{T}}$ includes the mean μ_q and the variance σ_q^2 of the signal on the $(q+1)$-th segment ($q \in [\![0\,,Q]\!]$).

Straightforward computations lead to the following explicit expressions

• for $\boldsymbol{F}(\boldsymbol{\eta}_q)$, $q \in [\![0\,,Q]\!]$,

$$\boldsymbol{F}(\boldsymbol{\eta}_q) = \operatorname{diag}\left(\frac{1}{\sigma_q^2}, \frac{1}{2(\sigma_q^2)^2}\right), \tag{36}$$

• for $\rho_q(s)$, $q \in [\![1\,,Q]\!]$, $s \in \,]0\,,1[$,

$$\rho_q(s) = \sqrt{\frac{\left(\nu^{\mathsf{v}}_{q-1,q}\right)^s}{s\,\nu^{\mathsf{v}}_{q-1,q}+1-s}}\exp\left\{-\frac{s(1-s)\,\nu^{\mathsf{m}}_{q-1,q}}{2(s\,\nu^{\mathsf{v}}_{q-1,q}+1-s)}\right\}, \tag{37}$$

in which we have defined the following two quantities, for $(q,q') \in [\![0\,,Q]\!]^2$:

$$\nu^{\mathsf{m}}_{q,q'} \triangleq \frac{(\mu_{q'}-\mu_q)^2}{\sigma_q^2} \qquad \text{and} \qquad \nu^{\mathsf{v}}_{q,q'} \triangleq \frac{\sigma_{q'}^2}{\sigma_q^2}, \tag{38}$$

which correspond to the squares of the so-called "amount of change" (also sometimes referred to as "magnitude of change" or "signal-to-noise ratio") for the mean and variance, respectively.

• for $R_q(s,s')$, $q \in [\![1\,,Q-1]\!]$ and $(s,s') \in \,]0\,,1[^2$, after tedious computations

$$\begin{aligned} R_q(s,s') &= \sqrt{\frac{s\,\nu^{\mathsf{v}}_{q-1,q+1}+(s'-s)\,\nu^{\mathsf{v}}_{q,q+1}+1-s'}{(s\,\nu^{\mathsf{v}}_{q-1,q}+1-s)(s'\,\nu^{\mathsf{v}}_{q,q+1}+1-s')}} \\ &\times \exp\left\{-\frac{s(1-s')}{2(s\,\nu^{\mathsf{v}}_{q-1,q+1}+(s'-s)\,\nu^{\mathsf{v}}_{q,q+1}+1-s')}\right. \\ &\qquad \times\left(\frac{s\,\nu^{\mathsf{v}}_{q-1,q+1}+1-s}{s\,\nu^{\mathsf{v}}_{q-1,q}+1-s}\,\nu^{\mathsf{m}}_{q-1,q}\right. \\ &\left.\left.+\frac{s'\,\nu^{\mathsf{v}}_{q-1,q+1}+1-s'}{s'\,\nu^{\mathsf{v}}_{q,q+1}+1-s'}\,\nu^{\mathsf{m}}_{q,q+1}-\nu^{\mathsf{m}}_{q-1,q+1}\right)\right\}, \end{aligned} \tag{39}$$

• for $\varphi_{\mu_{\tilde{q}},q}(s)$, $q \in [\![0\,,Q]\!]$, $\ell \in [\![1\,,L]\!]$, $\tilde{q} \in \{q-1,q\}$, $s \in \,]0\,,1[$, we obtain:

$$\varphi_{\mu_{\tilde{q}},q}(s) = \frac{\rho_q(s)}{\sigma_{\tilde{q}}^2}\left(\frac{s\mu_{q-1}\,\nu^{\mathsf{v}}_{q-1,q}+(1-s)\mu_q}{s\,\nu^{\mathsf{v}}_{q-1,q}+1-s}-\mu_{\tilde{q}}\right), \tag{40}$$

and

$$\begin{aligned} \varphi_{\sigma^2_{\tilde{q}},q}(s) &= \frac{\rho_q(s)}{2\sigma_{\tilde{q}}^2}\left(\frac{\left(s\mu_{q-1}\,\nu^{\mathsf{v}}_{q-1,q}+(1-s)\mu_q\right)^2}{\sigma_{\tilde{q}}^2(s\,\nu^{\mathsf{v}}_{q-1,q}+1-s)^2}\right. \\ &\left.+\frac{\sigma_q^2-2\mu_{\tilde{q}}(s\mu_{q-1}\,\nu^{\mathsf{v}}_{q-1,q}+(1-s)\mu_q)}{\sigma_{\tilde{q}}^2(s\,\nu^{\mathsf{v}}_{q-1,q}+1-s)}+\frac{\mu_{\tilde{q}}^2}{\sigma_{\tilde{q}}^2}-1\right). \end{aligned} \tag{41}$$

Using these expressions and plugging them into the appropriate equations from Section IV-A lead to the HB for a Gaussian signal submitted to Q abrupt changes.

2) Poisson case: The case of Poisson observations is also of interest in a number of signal processing applications, for instance for the segmentation of (possibly multivariate) astronomical time series [2], [43].

Let us assume that the observations are modeled according to (1), where the distribution on each segment is Poisson, i.e., for $q \in [\![0\,,Q]\!]$ and $n \in [\![t_q+1\,,t_{q+1}]\!]$ (with $t_0 \triangleq 0$ and $t_{Q+1} \triangleq N$), we have $x_n \sim \mathcal{P}(\lambda_q)$, or equivalently, $f(x_n\,;\lambda_q) = \Pr(X = x_n\,;\lambda_q) = \exp\{-\lambda_q\}\lambda_q^{x_n}/x_n!$. In this case, we have $\eta_q \triangleq \lambda_q$ in the $(q+1)$-th segment, which is a scalar parameter($L = 1$). Similarly to the case of Gaussian observations and after some computations, we obtain the following explicit expressions:

• For $F(\lambda_q)$, $q \in [\![0\,,Q]\!]$, $F(\lambda_q) = \frac{1}{\lambda_q}$,

• for $\rho_q(s)$, $q \in [\![1\,,Q]\!]$, $s \in \,]0\,,1[$,

$$\rho_q(s) = \exp\left\{-s\lambda_{q-1}-(1-s)\lambda_q+\lambda_{q-1}^s\lambda_q^{1-s}\right\}, \tag{42}$$

• for $R_q(s,s')$, $q \in [\![1\,,Q-1]\!]$ and $(s,s') \in \,]0\,,1[^2$, tedious computations lead to

$$\begin{aligned} R_q(s,s') = \exp\Bigg\{-\lambda_q\Bigg[1-\left(\frac{\lambda_{q-1}}{\lambda_q}\right)^s-\left(\frac{\lambda_{q+1}}{\lambda_q}\right)^{1-s'} \\ +\left(\frac{\lambda_{q-1}}{\lambda_q}\right)^s\left(\frac{\lambda_{q+1}}{\lambda_q}\right)^{1-s'}\Bigg]\Bigg\}, \end{aligned} \tag{43}$$

• for $\varphi_{\lambda_{\tilde{q}},q}(s)$, $q \in [\![0\,,Q]\!]$, $\ell \in [\![1\,,L]\!]$, $\tilde{q} \in \{q-1,q\}$, $s \in \,]0\,,1[$, tedious computations yield

$$\varphi_{\lambda_{\tilde{q}},q}(s) = \rho_q(s)\left(\frac{\lambda_{q-1}^s\lambda_q^{1-s}}{\lambda_{\tilde{q}}}-1\right). \tag{44}$$

We finally obtain the HB for a Poisson distributed signal that includes Q change-points by plugging these last expressions into the equations (17), (19), (20), (24), (25), (32), (33) and (34) from Section IV-A, and by applying the procedure described in Section III-C, which leads to the tightest bound.

V. Numerical results

This section presents some simulation results that enable us to assess the tightness of the proposed bound. It is compared in terms of global mean square error (GMSE) with the so-called ML-MAP estimator, for the distributions discussed in Section IV-B. All the cases discussed in this section were simulated with $N = 100$ observations, $Q = 2$ or 3 changes, $D = D_{\max}$ and $d = 6$ (except for Fig. 3), and the GMSE of the ML-MAP estimator was obtained by computing the empirical MSE through 1000 Monte-Carlo simulations.

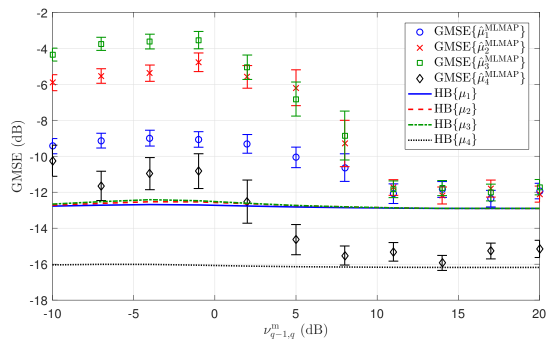

Figure 1. Empirical GMSEs and HBs for the mean estimates on each segment, for $Q = 3$ changes in the mean of $N = 100$ Gaussian observations.

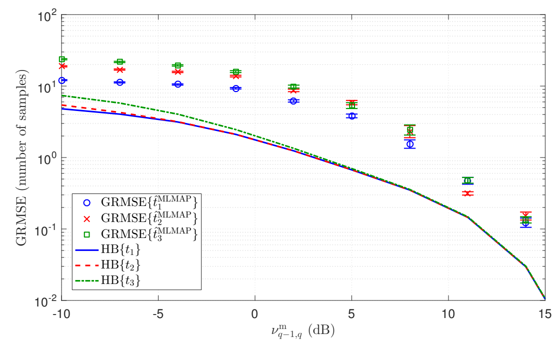

Figure 2. Empirical GRMSEs and HBs for the change-point estimates, for $Q = 3$ changes in the mean of $N = 100$ Gaussian observations.

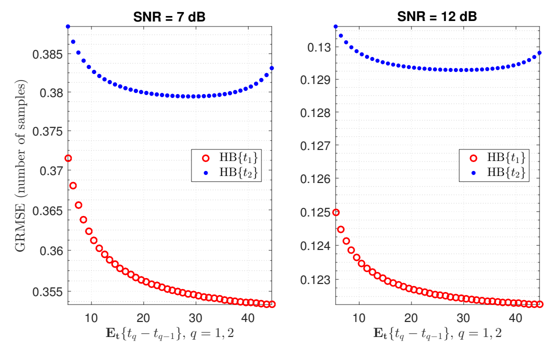

Figure 3. HBs for the change-point location estimates on each segment, for $Q = 2$ changes in the mean of Gaussian observations ($N = 100$), as a function of the prior average length of the segments $[\![t_{q-1}\,, t_q]\!]$, $q = 1, \ldots, 3$ (with the usual conventions $t_0 \triangleq 0$ and $t_3 \triangleq N$).

At each Monte-Carlo run, the $Q = 2$ or 3 changes were generated according to the prior distribution (2). The figures also display the "± 2 standard deviations" error bars for the estimated GMSE.

A. ML-MAP estimator

The ML-MAP estimator can be used when some elements of the parameter vector are deterministic and the others are random variables [20, p. 12], [21]. In our case, it is defined as follows:

$$(\hat{\boldsymbol{\eta}}^{\mathsf{MLMAP}}, \hat{\boldsymbol{t}}^{\mathsf{MLMAP}}) \triangleq \arg\max_{\boldsymbol{\eta}, \boldsymbol{t}} \ln f(\boldsymbol{x}, \boldsymbol{t}\,; \boldsymbol{\eta}). \tag{45}$$

Looking at (2), we can see that, as long as $\boldsymbol{t}$ belongs to its support $\mathscr{T}'$, the expression of the prior function $\pi(\boldsymbol{t})$ does not explicitly depend on $\boldsymbol{t}$. Consequently, the joint likelihood in (45) can be replaced with the classical likelihood $f(\boldsymbol{x} \,|\, \boldsymbol{t}\,; \boldsymbol{\eta})$. In addition, in the cases we study in the following sections (i.e., Gaussian and Poisson distributions), the maximization of the log likelihood w.r.t. $\boldsymbol{\eta}$, for a given $\boldsymbol{t}$, is not difficult and results in classical expressions of the empirical mean and/or variance for $\hat{\boldsymbol{\eta}}(\boldsymbol{x}\,; \boldsymbol{t})$. Hence, it is possible to get rid of the dependence on $\boldsymbol{\eta}$ in (45), so that we obtain $\hat{\boldsymbol{t}}^{\mathsf{MLMAP}}$ by using $\hat{\boldsymbol{t}}^{\mathsf{MLMAP}} = \arg\max_{\boldsymbol{t}} \ln f(\boldsymbol{x} \,|\, \boldsymbol{t}\,; \hat{\boldsymbol{\eta}}(\boldsymbol{x}\,; \boldsymbol{t}))$leading to $\hat{\boldsymbol{\eta}}^{\mathsf{MLMAP}} = \hat{\boldsymbol{\eta}}(\boldsymbol{x}\,; \hat{\boldsymbol{t}}^{\mathsf{MLMAP}})$.

B. Changes in the mean of a Gaussian distribution

In this classical change-point estimation problem, the parameter vector $\boldsymbol{\eta}$ contains the means μ_q, $q = 0, \ldots, Q$ of each segment, and possibly the corresponding variances σ_q^2, $q = 0, \ldots, Q$ (even if they remain constant), depending on whether they are assumed to be known or not. Extensive simulations have shown that the bound obtained by setting $s_q = 0.5$, $\forall q$ is tighter than with other values of s_q. Such simulations are not reported here due to the lack of space, but are very similar to those presented in [18]. Figs. 1 and 2 display the GMSE and the associated HB for the means $\mu_1, \ldots, \mu_4$ of each segment, and for the change-point locations t_1, t_2, t_3 as functions of the squared amount of change ν, respectively. These quantities correspond to the diagonal elements of the matrices in (4). Variations of the variance estimates and their HBs are very similar to Fig. 1. Due to the lack of space, such graphs are not reported here. Note that the global root mean square error (GRMSE) of the estimated change-point locations was plotted instead of the GMSE, for a more relevant assessment of the gap with the bound. The X-axis corresponds to $\nu^{\mathsf{m}}_{q,q'}$ defined in (38), quantifying the amount of change. More precisely, the Q changes generated for this experiment have all an equal amount of change, i.e., $\nu^{\mathsf{m}}_{0,1} = \nu^{\mathsf{m}}_{1,2} = \nu^{\mathsf{m}}_{2,3} = \nu^{\mathsf{m}}$ for a given ν^{m} such that $\mu_q = \mu_{q-1} + (-1)^{q-1}\nu^{\mathsf{m}}$ for $q = 1, \ldots, Q$, according to (38).

Fig. 1 clearly shows a threshold effect for the ML-MAP of the mean estimates, whose GMSEs move away from the HB for $\nu < 15\,\text{dB}$. The threshold is lower regarding the last segment mean, around 10 dB. The HB renders this behavior very slightly, as can be seen from the tiny bulge in the shape of the bound, for $0\,\text{dB} \leqslant \nu \leqslant 5\,\text{dB}$. For $\nu \geqslant 15\,\text{dB}$, the GMSEs and HBs become much closer one to the other.

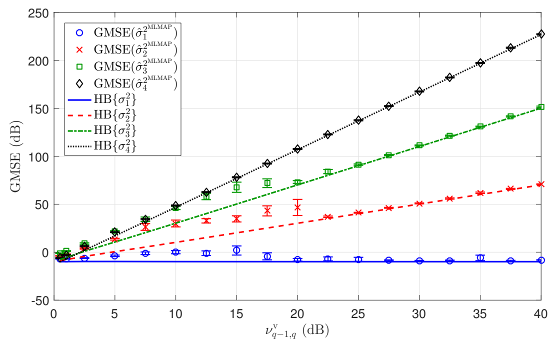

Figure 4. Empirical GMSEs and HBs for the variance estimates on each segment, for $Q = 3$ changes in the variance of Gaussian observations ($N = 100$).

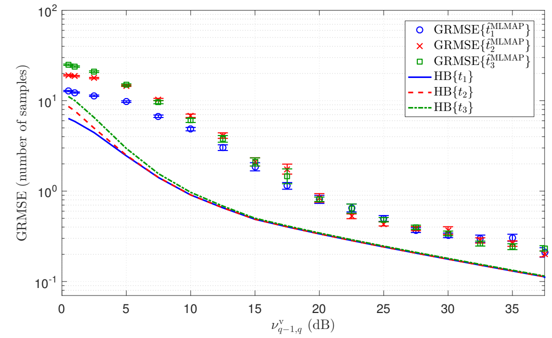

Figure 5. Empirical GRMSEs and HBs for the change-point location estimates, for $Q = 3$ changes in the variance of Gaussian observations ($N = 100$).

The small gap remaining in between, at high amounts of change, comes from the fact that the "CRB part" in the HB actually corresponds, as already mentioned, to a so-called *modified* CRB (MCRB), in the sense of [34]. Let us recall that the MCRB cannot be expected to be as tight as the classical CRB or the true Bayesian CRB, since the Fisher information is averaged over all the possible values of the nuisance parameter – in this case $\boldsymbol{t}$.

Regarding the estimation performance of change-point locations displayed in Fig. 2, both shapes of the GRMSEs and HBs highlight the existence of a non-information zone at low amounts of change, noticeable from the curve flatness. In this area, the HB shows that the early changes are better estimated than the later ones. This is an obvious effect of the prior support, which is larger for the late changes. Outside this non-information zone, the difference between GMSE and HB decreases: it is of the order of 9 or 10 samples for $\nu^{\mathsf{m}} = 0\,\text{dB}$ while it is lower than 2 samples for $\nu^{\mathsf{m}} > 10\,\text{dB}$, and lower than 0.1 samples for $\nu^{\mathsf{m}} = 15\,\text{dB}$. Curves cannot be shown for higher amounts of change because both empirical GRMSE and HB tend drastically to zero, due to the discrete nature of the change-point locations.

It is worth mentioning that the gap between the change-point location estimates and the bound is due to the discrete nature of these parameters. Indeed, discrete parameter estimation is not the most usual estimation framework, and the classical convergence theorems (regarding the MLE for instance) no longer apply (see [44] for general considerations on discrete parameter estimation).

In Fig. 3, we consider the case of $Q = 2$ changes, set $D = D_{\mathsf{max}}$ and $\Delta = 10$, and d varies. The HBs for both changes t_1 and t_2 are displayed as a function of $\mathbb{E}_{\boldsymbol{t}}\{t_q - t_{q-1}\} = (d+D)/2$, $q = 1, \dots, Q$ (with the usual conventions $t_0 \triangleq 0$ and $t_{Q+1} \triangleq N$), that is the prior average length of the segments $[\![t_{q-1}\,, t_q]\!]$. As $\mathbb{E}_{\boldsymbol{t}}\{t_q - t_{q-1}\}$ increases, the distance between the two changes t_1 and t_2 increases on average, according to the prior (2). We can see that the HBs w.r.t. both changes t_1 and t_2 are higher for small values of $\mathbb{E}_{\boldsymbol{t}}\{t_q - t_{q-1}\}$. This observation is consistent with the intuition that the closer the consecutive changes (or equivalently the smaller $\mathbb{E}_{\boldsymbol{t}}\{t_q - t_{q-1}\}$), the more difficult the estimation. For high values of $\mathbb{E}_{\boldsymbol{t}}\{t_q - t_{q-1}\}$, the HB w.r.t. t_1 decreases, since the number of observations used to infer it (on the first and the second segment) grows on average. The behavior differs regarding the second change t_2, whose HB increases again for high values of $\mathbb{E}_{\boldsymbol{t}}\{t_q - t_{q-1}\}$ (say $\mathbb{E}_{\boldsymbol{t}}\{t_q - t_{q-1}\} \geqslant 25$). It is due to the fact that an increase in $\mathbb{E}_{\boldsymbol{t}}\{t_2 - t_1\}$ implies a decrease in the length of the last segment $[\![t_2\,, N]\!]$, thus fewer observations of the last segment are used to infer the position of the last change, making the estimation harder.

C. Changes in the variance of a Gaussian distribution

This case is treated similarly as the previous one, with the only difference that instead of changes in the mean, we study changes in the variance of the observations. Here, the parameter vector $\boldsymbol{\eta}$ includes at least the variances on each segment σ_q^2, $q = 0, \dots, Q$, and possibly the means μ_q, $q = 0, \dots, Q$ if they are unknown. The proposed simulations were obtained for unknown means, all equal to 1. Here, the amount of change corresponds to $\nu^{\mathsf{v}}_{q,q'}$ defined in (38): it is the amount of change in terms of variance such that $\sigma_q^2 = \nu^{\mathsf{v}} \times \sigma_{q-1}^2$, for $q = 1, \dots, Q$, and for a given ν^{v}. The estimated MSEs and the corresponding bounds for the variances and change locations are displayed in Figs. 4-5. Note that results for the mean estimates are not shown here but are very similar to those obtained for variance estimates in Fig. 4. The same remarks as those made for mean changes (see previous subsection) are valid in the present case, with the following slight differences: the non-information zone, regarding the variance estimates $\widehat{\sigma}_2^{2^{\mathsf{MLMAP}}}$ and $\widehat{\sigma}_3^{2^{\mathsf{MLMAP}}}$, ranges from 0 dB to 25 dB, while it ranges from 0 dB to 15 dB regarding $\widehat{\sigma}_1^{2^{\mathsf{MLMAP}}}$. For $\widehat{\sigma}_4^{2^{\mathsf{MLMAP}}}$, one cannot distinguish such non-information zone. With regard to the change-point

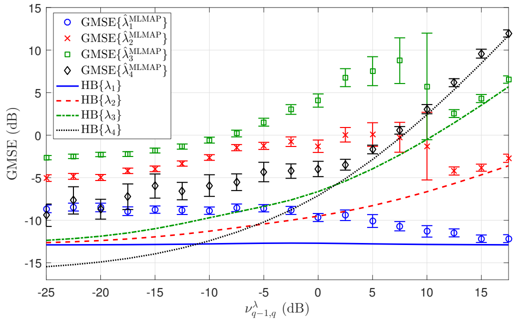

Figure 6. Empirical GMSEs and HBs for the mean rate estimates on each segment, for $Q = 3$ changes in the mean rate of Poisson observations ($N = 100$).

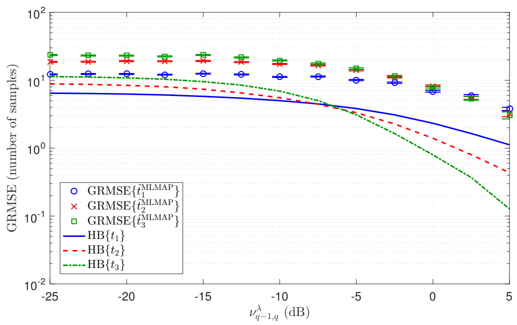

Figure 7. Empirical GRMSEs and HBs for the change-point locations estimates, for $Q = 3$ changes in the mean rate of Poisson observations ($N = 100$).

location estimates, the gap between the GRMSEs and the HBs becomes lower than 0.1 samples for $\nu^{\mathsf{v}} \geqslant 25$ dB.

D. Changes in the mean rate of a Poisson distribution

In the case of changes in the mean rate of a Poisson distribution, the parameter vector $\boldsymbol{\eta}$ includes the mean rates on each segment λ_q, $q = 0, \ldots, Q$. The results regarding mean rate and change-point estimates are displayed in Figs. 6-7. Note that the amount of change in these figures corresponds to the following definition (also used in [15], [16], [19]) $\nu^{\mathcal{P}}_{q,q'} \triangleq (\lambda_{q'} - \lambda_q)^2 / \lambda_q^2$. Hence, for a given $\nu^{\mathcal{P}}$, we have set $\lambda_q = \lambda_{q-1}(1 + \sqrt{\nu^{\mathcal{P}}})$, $q = 1, \ldots, Q$. The non-information zone for the mean rate estimates corresponds to amounts of change lower than 5 dB (see Fig. 6). According to Fig. 7, the non-information zone for the change location estimates corresponds to amounts of change lower than -15 dB. Other comments made in the Gaussian case are valid here.

VI. Conclusion

In this paper, we derived closed-form expressions of lower bounds on the MSE for parameter estimates of signals subjected to multiple change-points. The problem is challenging in that the unknown parameter vector contains both continuous and discrete parameters. The proposed approach consists in deriving the HB, which corresponds to the combination of the classical Cramér-Rao bound, for the (continuous) noise distribution parameters, with the Weiss-Weinstein bound, for the (discrete) change location parameters. Numerical simulations allowed the tightness of the bound to be assessed in two interesting scenarios: 1) mean value and variance changes for Gaussian random variables, and 2) mean rate changes for Poisson random variables. The proposed bound can be used as a reference to compare the performance of various parameter estimators for signals subjected to multiple change-points. More importantly, even if the bound is not rigorously attained, it provides a good approximation of the asymptotic mean square error of the unknown model parameters, i.e., the distribution parameters in each segment and the change-point locations. The expressions of the proposed bounds showed that the performance of parameter estimators in the presence of change-points is mainly driven by the signal-to-noise ratio of each change (corresponding to the amount of change between the different segments) and by the length of these segments. As a consequence, these bounds can be used to assess the difficulty of some change-point estimation problem.

Some aspects that have not been considered in this paper could be the subjects of future works, such as the influence of the sampling period, an unknown number of changes Q, or non i.i.d. observations.

Appendix A
Derivation of the matrix block $\boldsymbol{V}_{11}$

Since $f(\boldsymbol{x}, \boldsymbol{t}\,; \boldsymbol{\eta}) = f(\boldsymbol{x} \,|\, \boldsymbol{t}\,; \boldsymbol{\eta})\, \pi(\boldsymbol{t})$, we have

$$\frac{\partial^2 \ln f(\boldsymbol{x}, \boldsymbol{t}\,; \boldsymbol{\eta})}{\partial \eta_{q,\ell} \partial \eta_{q',\ell'}} = \frac{\partial^2 \ln f(\boldsymbol{x} \,|\, \boldsymbol{t}\,; \boldsymbol{\eta})}{\partial \eta_{q,\ell} \partial \eta_{q',\ell'}},$$

due to the independence between the assumed change-point prior and $\boldsymbol{\eta}$. From (3), the log-likelihood is deduced as $\ln f(\boldsymbol{x} \,|\, \boldsymbol{t}\,; \boldsymbol{\eta}) = \sum_{q=0}^{Q} \sum_{n=t_q+1}^{t_{q+1}} \ln f(x_n\,; \boldsymbol{\eta}_q)$, thus its first derivative w.r.t. $\eta_{q,\ell}$, with $(q, q') \in [\![0\,, Q]\!]^2$ and $(\ell, \ell') \in [\![1\,, L]\!]^2$, can be written as

$$\frac{\partial \ln f(\boldsymbol{x} \,|\, \boldsymbol{t}\,; \boldsymbol{\eta})}{\partial \eta_{q,\ell}} = \sum_{n=t_q+1}^{t_{q+1}} \frac{\partial \ln f(x_n\,; \boldsymbol{\eta}_q)}{\partial \eta_{q,\ell}}. \tag{A.46}$$

Hence, it can directly be noticed that

$$\forall q' \neq q, \ \frac{\partial^2 \ln f(\boldsymbol{x} \,|\, \boldsymbol{t}\,; \boldsymbol{\eta})}{\partial \eta_{q,\ell} \partial \eta_{q',\ell'}} = 0, \tag{A.47}$$

which leads to the block-diagonal structure of $\boldsymbol{V}_{11}$. We now assume that $q' = q$. By differentiating (A.46) w.r.t. $\eta_{q_{\ell'}}$, and taking the expectation w.r.t. the joint likelihood between $\boldsymbol{x}$ and $\boldsymbol{t}$, we obtain, by setting $p = Lq + \ell$ and $p' = Lq + \ell'$,

$$[\boldsymbol{V}_{11}]_{p,p'} = -\sum_{\boldsymbol{t}\in\mathbb{Z}^Q} \pi(\boldsymbol{t}) \sum_{n=t_q+1}^{t_{q+1}} \int_\Omega \left[\prod_{\substack{i=0\\ i\neq q}}^{Q} \prod_{m=t_i+1}^{t_{i+1}} f(x_m\,;\boldsymbol{\eta}_i) \right] \times \left[\left(\prod_{\substack{m=t_q+1\\ m\neq n}}^{t_{q+1}} f(x_m\,;\boldsymbol{\eta}_q) \right) \frac{\partial^2 \ln f(x_n\,;\boldsymbol{\eta}_q)}{\partial \eta_{q,\ell} \partial \eta_{q,\ell'}} f(x_n\,;\boldsymbol{\eta}_q) \right] \mathrm{d}\boldsymbol{x}, \tag{A.48}$$

by splitting the double product. Since variables are separated, we can switch product and integral signs, leading to

$$[\boldsymbol{V}_{11}]_{p,p'} = -\frac{1}{\Delta^Q} \sum_{\substack{t_j=t_{j-1}+d\\ j\in[\![1,Q]\!]}}^{t_{j-1}+D} \sum_{n=t_q+1}^{t_{q+1}} \int_{\Omega'} \frac{\partial^2 \ln f(x\,;\boldsymbol{\eta}_q)}{\partial \eta_{q,\ell} \partial \eta_{q,\ell'}} f(x\,;\boldsymbol{\eta}_q)\,\mathrm{d}x, \tag{A.49}$$

by removing the integrals that equal 1. Note that the last integral actually neither depends on n, nor on $\boldsymbol{t}$, and that it corresponds to the (ℓ, ℓ') element of the Fisher information matrix for the parameter vector $\boldsymbol{\eta}_q$, for one observation in the q-th segment, as defined in (18). Hence, we obtain

$$[\boldsymbol{V}_{11}]_{p,p'} = \begin{cases} \frac{d+D}{2} \left[\boldsymbol{F}(\boldsymbol{\eta}_q)\right]_{\ell,\ell'} & \text{if } q < Q, \\ \left(N - \frac{Q(d+D)}{2}\right) \left[\boldsymbol{F}(\boldsymbol{\eta}_Q)\right]_{\ell,\ell'} & \text{if } q = Q. \end{cases} \tag{A.50}$$

Appendix B
Derivation of the Matrix Block $\boldsymbol{V}_{12}$

In this appendix, we derive the expression of the left-hand side of (31). For this purpose, letting $\tilde{q} \in [\![0\,,Q]\!]$, $\ell \in [\![1\,,L]\!]$ and $q \in [\![1\,,Q]\!]$, we first derive the quantity

$$\mathbb{E}_{\boldsymbol{x},\boldsymbol{t};\boldsymbol{\eta}} \left\{ \frac{\partial \ln f(\boldsymbol{x},\boldsymbol{t}\,;\boldsymbol{\eta})}{\partial \eta_{\tilde{q},\ell}} \frac{f^{s_q}(\boldsymbol{x},\boldsymbol{t}+\boldsymbol{h}_q\,;\boldsymbol{\eta})}{f^{s_q}(\boldsymbol{x},\boldsymbol{t}\,;\boldsymbol{\eta})} \right\} = \sum_{\boldsymbol{t}\in\mathbb{Z}^Q} \pi^{s_q}(\boldsymbol{t}+\boldsymbol{h}_q)\, \pi^{1-s_q}(\boldsymbol{t}) \int_\Omega \sum_{n=t_{\tilde{q}}+1}^{t_{\tilde{q}+1}} \frac{\partial \ln f(x_n\,;\boldsymbol{\eta}_{\tilde{q}})}{\partial \eta_{\tilde{q},\ell}} \times f^{s_q}(\boldsymbol{x}\,|\,\boldsymbol{t}+\boldsymbol{h}_q\,;\boldsymbol{\eta}) f^{1-s_q}(\boldsymbol{x}\,|\,\boldsymbol{t}\,;\boldsymbol{\eta})\,\mathrm{d}\boldsymbol{x}. \tag{B.51}$$

On the one hand, from (3) and the form of $\boldsymbol{h}_q$, we can write $f(\boldsymbol{x}\,|\,\boldsymbol{t}+\boldsymbol{h}_q\,;\boldsymbol{\eta}) = \prod_{i=0}^{Q} \prod_{n=t_i+\delta_{i,q}h_q+1}^{t_{i+1}+\delta_{i,q-1}h_q} f(x_n\,;\boldsymbol{\eta}_i)$, where $\delta_{i,q}$ denotes the Krönecker delta (i.e., it equals 1 if $i = q$, zero otherwise). Consequently, for any sign of h_q,

$$f^{s_q}(\boldsymbol{x}\,|\,\boldsymbol{t}+\boldsymbol{h}_q\,;\boldsymbol{\eta})\, f^{1-s_q}(\boldsymbol{x}\,|\,\boldsymbol{t}\,;\boldsymbol{\eta}) = \left(\prod_{i=0}^{Q} \prod_{m=t_i+\delta_{i,q}(h_q)^++1}^{t_{i+1}-\delta_{i,q-1}(-h_q)^+} f(x_m\,;\boldsymbol{\eta}_i) \right) \times \prod_{m=t_q-(-h_q)^++1}^{t_q+(h_q)^+} f^{\varepsilon_{h_q}(s_q)}(x_m\,;\boldsymbol{\eta}_{q-1})\, f^{1-\varepsilon_{h_q}(s_q)}(x_m\,;\boldsymbol{\eta}_q), \tag{B.52}$$

where $\varepsilon_{h_q}(s_q)$ is defined in (23), and for $x \in \mathbb{R}$, $(x)^+ \triangleq \max(x, 0)$.

On the other hand, from (2), we have

$$\pi^{s_q}(\boldsymbol{t}+\boldsymbol{h}_q)\, \pi^{1-s_q}(\boldsymbol{t}) = \frac{1}{\Delta^Q} \mathbb{1}_{\mathcal{T}'}(\boldsymbol{t}+\boldsymbol{h}_q)\, \mathbb{1}_{\mathcal{T}'}(\boldsymbol{t}) = \frac{1}{\Delta^Q} \mathbb{1}_{\mathcal{S}}(\boldsymbol{t}) \tag{B.53}$$

where $\mathcal{S} \triangleq (\mathcal{T}' - \boldsymbol{h}_q) \cap \mathcal{T}'$, with $\mathcal{T}'$ denoting the support of the prior distribution $\pi(\boldsymbol{t})$, and we use the abuse of notation $\mathcal{T}' - \boldsymbol{h}_q$ to refer to the translated set $\left\{\boldsymbol{t} \in \mathbb{Z}^Q \,\middle|\, \boldsymbol{t}+\boldsymbol{h}_q \in \mathcal{T}'\right\}$.

From (B.52) and (B.53), we can continue the derivation leading to (B.51), by considering the following three cases:
1) $\tilde{q} \neq q-1$ and $\tilde{q} \neq q$, case referred to as "ULT" (for "upper and lower triangles");
2) $\tilde{q} = q-1$, case referred to as "D1" (for "1st diagonal");
3) $\tilde{q} = q$, case referred to as "D2" (for "2nd diagonal").
These three cases are studied in the following sections.

A. Case ULT ($\tilde{q} \neq q-1$ and $\tilde{q} \neq q$)

By plugging (B.52) and (B.53) into (B.51), and using the same kind of manipulations as for (A.49), we obtain

$$\mathbb{E}_{\boldsymbol{x},\boldsymbol{t};\boldsymbol{\eta}} \left\{ \frac{\partial \ln f(\boldsymbol{x},\boldsymbol{t}\,;\boldsymbol{\eta})}{\partial \eta_{\tilde{q},\ell}} \frac{f^{s_q}(\boldsymbol{x},\boldsymbol{t}+\boldsymbol{h}_q\,;\boldsymbol{\eta})}{f^{s_q}(\boldsymbol{x},\boldsymbol{t}\,;\boldsymbol{\eta})} \right\} = \frac{1}{\Delta^Q} \sum_{\boldsymbol{t}\in\mathbb{Z}^Q} \mathbb{1}_{\mathcal{S}}(\boldsymbol{t}) \times \sum_{n=t_{\tilde{q}}+1}^{t_{\tilde{q}+1}} \left[\prod_{\substack{i=0\\ i\neq\tilde{q}}}^{Q} \prod_{m=t_i+\delta_{i,q}(h_q)^++1}^{t_{i+1}-\delta_{i,q-1}(-h_q)^+} \int_{\Omega'} f(x_m\,;\boldsymbol{\eta}_i)\,\mathrm{d}x_m \times \left(\prod_{\substack{m=t_{\tilde{q}}+1\\ m\neq n}}^{t_{\tilde{q}+1}} \int_{\Omega'} f(x_m\,;\boldsymbol{\eta}_{\tilde{q}})\,\mathrm{d}x_m \right) \int_{\Omega'} \frac{\partial \ln f(x_n\,;\boldsymbol{\eta}_{\tilde{q}})}{\partial \eta_{\tilde{q},\ell}} f(x_n\,;\boldsymbol{\eta}_{\tilde{q}})\,\mathrm{d}x_n \times \prod_{m=t_q-(-h_q)^++1}^{t_q+(h_q)^+} \int_{\Omega'} f^{\varepsilon_{h_q}(s_q)}(x_m\,;\boldsymbol{\eta}_{q-1})\, f^{1-\varepsilon_{h_q}(s_q)}(x_m\,;\boldsymbol{\eta}_q)\,\mathrm{d}x_m \right]. \tag{B.54}$$

Yet, as a consequence of (8) and the independence of the observations, the integral $\int_{\Omega'} \frac{\partial \ln f(x_n;\boldsymbol{\eta}_{\tilde{q}})}{\partial \eta_{\tilde{q},\ell}} f(x_n\,;\boldsymbol{\eta}_{\tilde{q}})\,\mathrm{d}x_n$ in (B.54) is actually zero. Hence, for $\tilde{q} \in [\![0\,,Q]\!]$, $\ell \in [\![1\,,L]\!]$, $p = L\tilde{q} + \ell$, and $q \in [\![1\,,Q-1]\!]$ such that $\tilde{q} \neq q-1$ and $\tilde{q} \neq q$, we obtain $[\boldsymbol{V}_{12}]_{p,q} = 0$.

B. Case D1 ($\tilde{q} = q-1$)

For this case, the derivation of (B.51) depends upon whether $h_q > 0$ or $h_q < 0$.

- Let us first assume $h_q > 0$. This case is treated exactly as in Appendix B-A, i.e., one can rewrite (B.54) similarly just by appropriately replacing $\tilde{q}$ with $q-1$, and we also find

$$\mathbb{E}_{\boldsymbol{x},\boldsymbol{t};\boldsymbol{\eta}^\star} \left\{ \frac{\partial \ln f(\boldsymbol{x},\boldsymbol{t}\,;\boldsymbol{\eta})}{\partial \eta_{q-1,\ell}} \frac{f^{s_q}(\boldsymbol{x},\boldsymbol{t}+\boldsymbol{h}_q\,;\boldsymbol{\eta})}{f^{s_q}(\boldsymbol{x},\boldsymbol{t}\,;\boldsymbol{\eta})} \right\} = 0. \tag{B.55}$$

• We now assume $h_q < 0$ (for convenience, we replace "h_q" with "$-|h_q|$"). In equation (B.54), we have to split the discrete sum w.r.t. index n, so that rewriting (B.54) yields

$$\mathbb{E}_{\boldsymbol{x},\boldsymbol{t};\boldsymbol{\eta}}\left\{\frac{\partial \ln f(\boldsymbol{x},\boldsymbol{t};\boldsymbol{\eta})}{\partial \eta_{q-1,\ell}}\frac{f^{s_q}(\boldsymbol{x},\boldsymbol{t}+\boldsymbol{h}_q;\boldsymbol{\eta})}{f^{s_q}(\boldsymbol{x},\boldsymbol{t};\boldsymbol{\eta})}\right\}$$
$$= \frac{1}{\Delta^Q}\sum_{\boldsymbol{t}\in\mathbb{Z}^Q}\mathbb{1}_{\mathscr{S}}(\boldsymbol{t})\left[\sum_{n=t_{q-1}+1}^{t_q-|h_q|}\left(\prod_{\substack{i=0\\ i\neq q-1}}^{Q}\prod_{\substack{m=\\ t_i+1}}^{t_{i+1}}\int_{\Omega'} f(x_m;\boldsymbol{\eta}_i)\,\mathrm{d}x_m\right.\right.$$
$$\times\left(\prod_{\substack{m=t_{q-1}+1\\ m\neq n}}^{t_q-|h_q|}\int_{\Omega'} f(x_m;\boldsymbol{\eta}_{q-1})\,\mathrm{d}x_m\right)\int_{\Omega'}\frac{\partial \ln f(x_n;\boldsymbol{\eta}_{q-1})}{\partial \eta_{q-1,\ell}} \times f(x_n;\boldsymbol{\eta}_q)\,\mathrm{d}x_n$$
$$\left.\times\prod_{m=t_q-|h_q|+1}^{t_q}\int_{\Omega'} f^{1-s_q}(x_m;\boldsymbol{\eta}_{q-1})\,f^{s_q}(x_m;\boldsymbol{\eta}_q)\,\mathrm{d}x_m\right)$$
$$+\sum_{n=t_q-|h_q|+1}^{t_q}\left(\prod_{i=0}^{Q}\prod_{m=t_i+1}^{t_{i+1}-\delta_{i,q-1}|h_q|}\int_{\Omega'} f(x_m;\boldsymbol{\eta}_i)\,\mathrm{d}x_m\right.$$
$$\times\prod_{\substack{m=t_q-|h_q|+1\\ m\neq n}}^{t_q}\int_{\Omega'} f^{1-s_q}(x_m;\boldsymbol{\eta}_{q-1})\,f^{s_q}(x_m;\boldsymbol{\eta}_q)\,\mathrm{d}x_m$$
$$\left.\left.\times\int_{\Omega'}\frac{\partial \ln f(x_n;\boldsymbol{\eta}_{q-1})}{\partial \eta_{q-1,\ell}}f^{1-s_q}(x_n;\boldsymbol{\eta}_{q-1})\,f^{s_q}(x_n;\boldsymbol{\eta}_q)\,\mathrm{d}x_n\right)\right]. \tag{B.56}$$

Note that the sum indexed by $n \in [\![t_{q-1}+1\,,t_q-|h_q|]\!]$ in (B.56) is similar to (B.54) and consequently equals zero, still because of the regularity condition (8). The second sum, indexed by $n \in [\![t_q-|h_q|+1\,,t_q]\!]$, is nonzero and by using the definitions (22) and (35), and evaluating (B.56) in $\boldsymbol{\eta}^\star$, we obtain

$$\mathbb{E}_{\boldsymbol{x},\boldsymbol{t};\boldsymbol{\eta}^\star}\left\{\left.\frac{\partial \ln f(\boldsymbol{x},\boldsymbol{t};\boldsymbol{\eta})}{\partial \eta_{q-1,\ell}}\right|_{\boldsymbol{\eta}^\star}\frac{f^{s_q}(\boldsymbol{x},\boldsymbol{t}+\boldsymbol{h}_q;\boldsymbol{\eta}^\star)}{f^{s_q}(\boldsymbol{x},\boldsymbol{t};\boldsymbol{\eta}^\star)}\right\}$$
$$= \frac{|h_q|}{\Delta^Q}\,\rho_q^{|h_q|-1}(1-s_q)\,\varphi_{\eta_{q-1,\ell},q}(1-s_q)\sum_{\boldsymbol{t}\in\mathbb{Z}^Q}\mathbb{1}_{\mathscr{S}}(\boldsymbol{t}). \tag{B.57}$$

We complete the calculation by noticing that $\frac{1}{\Delta^Q}\sum_{\boldsymbol{t}\in\mathbb{Z}^Q}\mathbb{1}_{\mathscr{S}}(\boldsymbol{t}) = \frac{1}{\Delta^Q}\sum\cdots\sum_{\boldsymbol{t}\in\mathscr{S}}1 = u_{\mathsf{D}}(\Delta,h_q)$, thus

$$\mathbb{E}_{\boldsymbol{x},\boldsymbol{t};\boldsymbol{\eta}^\star}\left\{\left.\frac{\partial \ln f(\boldsymbol{x},\boldsymbol{t};\boldsymbol{\eta})}{\partial \eta_{q-1,\ell}}\right|_{\boldsymbol{\eta}^\star}\frac{f^{s_q}(\boldsymbol{x},\boldsymbol{t}+\boldsymbol{h}_q;\boldsymbol{\eta}^\star)}{f^{s_q}(\boldsymbol{x},\boldsymbol{t};\boldsymbol{\eta}^\star)}\right\}$$
$$= |h_q|\,u_{\mathsf{D}}(\Delta,h_q)\,\rho_q^{|h_q|-1}(1-s_q)\,\varphi_{\eta_{q-1,\ell},q}(1-s_q). \tag{B.58}$$

After plugging (B.55) and (B.58) into (31), we obtain (33).

C. Case D2 ($\tilde{q} = q$)

The derivation of (B.51) also depends upon whether $h_q > 0$ or $h_q < 0$. Actually, those two cases are treated very similarly as in the previous case D1, so that we finally obtain, as in Appendix B-B,

$$\mathbb{E}_{\boldsymbol{x},\boldsymbol{t};\boldsymbol{\eta}^\star}\left\{\left.\frac{\partial \ln f(\boldsymbol{x},\boldsymbol{t};\boldsymbol{\eta})}{\partial \eta_{q,\ell}}\right|_{\boldsymbol{\eta}^\star}\frac{f^{s_q}(\boldsymbol{x},\boldsymbol{t}+\boldsymbol{h}_q;\boldsymbol{\eta}^\star)}{f^{s_q}(\boldsymbol{x},\boldsymbol{t};\boldsymbol{\eta}^\star)}\right\}$$
$$= \begin{cases} |h_q|\,u_{\mathsf{D}}(\Delta,h_q)\,\rho_q^{|h_q|-1}(s_q)\,\varphi_{\eta_{q,\ell},q}(s_q) & \text{if } h_q > 0,\\ 0 & \text{if } h_q < 0.\end{cases} \tag{B.59}$$

After plugging (B.59) into (31), we obtain (34).

References

[1] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes, Theory and Application.* NJ: Prentice-Hall: Englewood Cliffs, 1993.

[2] J. D. Scargle, "Studies in astronomical time series analysis: V. Bayesian blocks, a new method to analyze structure in photon counting data," *Astrophysical Journal*, vol. 504, pp. 405–418, Sep. 1998.

[3] J. Chen and A. K. Gupta, *Parametric Statistical Change Point Analysis*. Birkhäuser Basel, 2000.

[4] J.-Y. Tourneret, M. Doisy, and M. Lavielle, "Bayesian off-line detection of multiple change-points corrupted by multiplicative noise: application to SAR image edge detection," *Signal Processing*, vol. 83, no. 9, pp. 1871–1887, Sep. 2003.

[5] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, 2nd ed., ser. Springer Texts in Statistics. New-York, NY, USA: Springer, Sep. 2003.

[6] D. V. Hinkley, "Inference about the change-point in a sequence of random variables," *Biometrika*, vol. 57, no. 1, pp. 1–18, 1970.

[7] S. B. Fotopoulos, S. K. Jandhyala, and E. Khapalova, "Exact asymptotic distribution of change-point MLE for change in the mean of Gaussian sequences," *Annals of Applied Statistics*, vol. 4, no. 2, pp. 1081–1104, Nov. 2010.

[8] H. L. V. Trees, K. L. Bell, and Z. Thian, *Detection Estimation and Modulation Theory, Part I: Detection, Estimation, and Filtering Theory*, 2nd ed. Wiley, 2013.

[9] E. W. Barankin, "Locally best unbiased estimates," *Annals of Mathematical Statistics*, vol. 20, no. 4, pp. 477–501, Dec. 1949.

[10] D. G. Chapman and H. Robbins, "Minimum variance estimation without regularity assumptions," *Annals of Mathematical Statistics*, vol. 22, no. 4, pp. 581–586, Dec. 1951.

[11] R. J. McAulay and E. M. Hofstetter, "Barankin bounds on parameter estimation," *IEEE Transactions on Information Theory*, vol. 17, no. 6, pp. 669–676, Nov. 1971.

[12] K. Todros and J. Tabrikian, "General classes of performance lower bounds for parameter estimation - part I: non-Bayesian bounds for unbiased estimators," *IEEE Transactions on Information Theory*, vol. 56, no. 10, pp. 5045–5063, Oct. 2010.

[13] E. Weinstein and A. J. Weiss, "A general class of lower bounds in parameter estimation," *IEEE Transactions on Information Theory*, vol. 34, no. 2, pp. 338–342, Mar. 1988.

[14] K. Todros and J. Tabrikian, "General classes of performance lower bounds for parameter estimation - part II: Bayesian bounds," *IEEE Transactions on Information Theory*, vol. 56, no. 10, pp. 5064–5082, Oct. 2010.

[15] A. Ferrari and J. Tourneret, "Barankin lower bound for change points in independent sequences," in *Proc. IEEE Workshop Statist. Signal Process.*, St. Louis, MO, USA, Sep. 2003, pp. 557–560.

[16] P. S. La Rosa, A. Renaux, A. Nehorai, and C. H. Muravchik, "Barankin-type lower bound on multiple change-point estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 11, pp. 5534–5549, Nov. 2010.

[17] L. Bacharach, A. Renaux, M. N. El Korso, and E. Chaumette, "Weiss-Weinstein bound for change-point estimation," in *Proc. of IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Cancún, Mexico, Dec. 2015, pp. 477–480.

[18] ——, "Weiss-Weinstein bound on multiple change-points estimation," *IEEE Transactions on Signal Processing*, vol. 65, no. 10, pp. 2686–2700, May 2017.

[19] L. Bacharach, M. N. El Korso, A. Renaux, and J.-Y. Tourneret, "A Bayesian Lower Bound for Parameter Estimation of Poisson Data Including Multiple Changes," in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, New Orleans, LA, USA, Mar. 2017, pp. 4486–4490.

[20] H. L. Van Trees and K. L. Bell, Eds., *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*. New-York, NY, USA: Wiley/IEEE Press, Sep. 2007.

[21] A. Yeredor, "The joint MAP-ML criterion and its relation to ML and to extended least-squares," *IEEE Transactions on Signal Processing*, vol. 48, no. 12, pp. 3484–3492, Dec. 2000.

[22] Y.-C. Yao, "Estimating the number of change-points via Schwarz' criterion," *Statistics and Probability Letters*, vol. 6, pp. 181–189, Feb. 1988.

[23] E. Lebarbier, "Detecting multiple change-points in the mean of gaussian process by model selection," *Signal Processing*, vol. 85, no. 4, pp. 717–736, Apr. 2005.

[24] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-33, no. 2, pp. 387–392, Apr. 1985.

[25] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood and the Cramér-Rao bound," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 5, pp. 720–741, May 1989.

[26] S. Bay, B. Geller, A. Renaux, J.-P. Barbot, and J.-M. Brossier, "On the hybrid Cramér-Rao bound and its application to dynamical phase estimation," *IEEE Signal Processing Letters*, vol. 15, pp. 453–456, 2008.

[27] I. Reuven and H. Messer, "A Barankin-type lower bound on the estimation error of a hybrid parameter vector," *IEEE Transactions on Information Theory*, vol. 43, no. 3, pp. 1084–1093, May 1997.

[28] K. Todros and J. Tabrikian, "Hybrid lower bound via compression of the sampled CLR function," in *Proc. IEEE Workshop Statist. Signal Process.*, Cardiff, Wales, UK, Aug. 2009, pp. 602–605.

[29] C. Ren, J. Galy, E. Chaumette, P. Larzabal, and A. Renaux, "Hybrid Barankin–Weiss-Weinstein bounds," *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 2064–2068, Nov. 2015.

[30] C. R. Rao, "Information and accuracy attainable in the estimation of statistical parameters," *Bulletin of the Calcutta Mathematical Society*, vol. 37, pp. 81–91, 1945.

[31] H. Cramér, *Mathematical Methods of Statistics*, ser. Princeton Mathematics. New-York: Princeton University Press, Sep. 1946, vol. 9.

[32] K. L. Bell and H. L. Van Trees, "Combined Cramér-Rao/Weiss-Weinstein bound for tracking target bearing," in *Proc. IEEE Workshop Sensor Array Multi-channel Process.*, Waltham, MA, USA, Jul. 2006, pp. 273–277.

[33] H. Messer, "The hybrid Cramér-Rao lower bound – from practice to theory," in *Proc. of IEEE Workshop on Sensor Array and Multi-channel Processing (SAM)*, Waltham, MA, USA, Jul. 2006, pp. 304–307.

[34] F. Gini, R. Reggiannini, and U. Mengali, "The modified Cramér-Rao bound in vector parameter estimation," *IEEE Transactions on Communications*, vol. 46, no. 1, pp. 52–60, Jan. 1998.

[35] W. Xu, "Performance bounds on matched-field methods for source localization and estimation of ocean environmental parameters," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA, Jun. 2001.

[36] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge, U.K.: Cambridge Univ. Press, 2004.

[37] M. Grant and S. Boyd, *CVX: Matlab software for disciplined convex programming*. [Online]. Available: *http://stanford.edu/boyd/cvx*, Web page and software, Dec. 2008.

[38] H. Chernoff and S. Zacks, "Estimating the current mean of a normal distribution which is subjected to changes in time," *Annals of Mathematical Statistics*, vol. 35, no. 3, pp. 999–1018, 1964.

[39] M. Lavielle and E. Lebarbier, "An application of MCMC methods for the multiple change-points problem," *Signal Processing*, vol. 81, pp. 39–53, Jan. 2001.

[40] 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, pp. 143–164, 2003.

[41] A. E. Raftery and V. E. Akman, "Bayesian analysis of a Poisson process with a change-point," *Biometrika*, vol. 73, no. 1, pp. 85–89, Apr. 1986.

[42] B. Jackson, J. S. Scargle, D. Barnes, S. Arabhi, A. Alt, P. Gioumousis, E. Gwin, P. Sangtrakulcharoen, L. Tan, and T. T. Tsai, "An algorithm for optimal partitioning of data on an interval," *IEEE Signal Processing Letters*, vol. 12, pp. 105–108, Feb. 2005.

[43] N. Dobigeon, J.-Y. Tourneret, and J. D. Scargle, "Joint segmentation of multivariate astronomical time series: Bayesian sampling with a hierarchical model," *IEEE Transactions on Signal Processing*, vol. 55, no. 2, pp. 414–423, Feb. 2007.

[44] C. Choirat and R. Seri, "Estimation in discrete parameter models," *Statistical Science*, vol. 27, no. 2, pp. 278–293, 2012.