Barankin-type lower bound on multiple change-point estimation

Patricio Salvatore La Rosa, Alexandre Renaux, Carlos H. Muravchik, Arye Nehorai

Barankin-type Lower Bound on Multiple Change-point Estimation

Patricio S. La Rosa, *Student Member, IEEE*, Alexandre Renaux, *Member, IEEE*, Carlos H. Muravchik, *Senior Member, IEEE*, and Arye Nehorai *Fellow, IEEE*

Abstract

We compute lower bounds on the mean-square error of multiple change-point estimation. In this context, the parameters are discrete and the Cramér-Rao bound is not applicable. Consequently, we focus on computing the Barankin bound (BB), the greatest lower bound on the covariance of any unbiased estimator, which is still valid for discrete parameters. In particular, we compute the multi-parameter version of the Hammersley-Chapman-Robbins, which is a Barankin-type lower bound. We first give the structure of the so-called Barankin information matrix (BIM) and derive a simplified form of the BB. We show that the particular case of two change points is fundamental to finding the inverse of this matrix. Several closed-form expressions of the elements of BIM are given for changes in the parameters of Gaussian and Poisson distributions. The computation of the BB requires finding the supremum of a finite set of positive definite matrices with respect to the *Loewner partial ordering*. Though, each matrix in this set of candidates is a lower bound on the covariance matrix of the estimator, the existence of a unique supremum *w.r.t.* to this set, i.e., the tightest bound, might not be guaranteed. To overcome this problem, we compute a suitable minimal-upper bound to this set given by the matrix associated with the Lowner-John Ellipsoid of the set of hyper-ellipsoids associated to the set of candidate lower-bound matrices. Finally, we present some numerical examples to compare the proposed approximated BB with the performance achieved by the maximum likelihood estimator.

This work was partially presented at the Second International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), December 12-14, 2007. St. Thomas, U.S. Virgin Islands.

Patricio S. La Rosa, and Arye Nehorai are with the Department of Electrical and Systems Engineering at Washington University in St. Louis, One Brookings Drive, St. Louis, MO 63130, USA. (e-mail: {pla,nehorai}@ese.wustl.edu).

A. Renaux is with the Laboratoire des Signaux et Systèmes, Laboratory of Signals and Systems, University Paris-Sud 11, 91192 Gif-sur-Yvette cedex, France (e-mail: renaux@lss.supelec.fr).

Carlos Muravchik is with the CIC-PBA and Departamento de Electrotecnia, Universidad Nacional de La Plata, La Plata, Argentina (e-mail: carlosm@ing.unlp.edu.ar).

This work was supported by the Department of Defense under the Air Force Office of Scientific Research MURI Grant FA9550-05-1-0443 and ONR Grant N000140810849.

I. INTRODUCTION

Estimation of changes in time series is an important and active research area with several applications, for example, fault detection, medical imaging, genetics, and econometrics. The literature is abundant concerning estimation algorithms for change-point estimation (see, *e.g.*, [1]–[3]). However, less work has been done concerning the ultimate performance of such algorithms in terms of mean-square error (MSE). Indeed, if an estimator is available, the evaluation of its performance depends on knowing whether it is optimal or if further improvement is still possible. Note that some other criteria of performance in the context of sequential detection of a change-point are available in the literature, see, *e.g.*, [4], [5] and references therein.

The classic way to analyze the performance of an estimator in terms of MSE is to compute the well-known Cramér-Rao bound (CRB) [6]. Unfortunately, for discrete time-measurement models the change-point location parameter is discrete; therefore the CRB, which is a function of the derivative of the likelihood of the observations w.r.t. the parameters, is not defined.

Several authors have proposed solutions to this problem. Indeed, in the change-point estimation framework, the CRB has already been studied using approximations (see, *e.g.*, [7]–[12]). Depending on the particular parametrization of the data likelihood, two main challenges have been addressed concerning the CRB computation on the change-point time index: (i) the discrete nature of the aforementioned parameter and (ii) the regularity conditions of the likelihood of the observation. The former implies that the parameter does not have a defined derivative because of its discrete nature [10], and the latter implies that the likelihood of the observations has to be smooth (details are given in [6] and [13]), which is not the case for signal parameters with sudden changes. To overcome the discrete nature of the change-point time index, a continuous parametrization has been proposed (see, *e.g.*, [12], [14]). To satisfy the regularity conditions of the data likelihood, the step-like function, which represents a change in parameter, is generally approximated by another function with smooth properties (*e.g.*, the so-called sigmoidal function introduced in [9] and [12] or a Heaviside function filtered by a Gaussian filter, as in [7]). This new function depends on parameters that have to be adjusted, and it tends to the step-like function when the appropriate values of these parameters are used. The main problem that appears when using this technique is that the CRB tends to zero when the approximate function tends to the step-like function [8], [12].

Moreover, it is noteworthy that these previous works concerning change-point estimation were always done in the framework of a single change point. To the best of our knowledge, performance bounds have

never been derived in the multiple change-point context. The latter is important in off-line estimation of change points where batch-data are available, for example, in biomedical applications, such as DNA sequence segmentation [15], rat EEG segmentation (see [3], Chapter 2), and uterine MMG contraction detection [16], and in signal segmentation in general such as speech segmentation [17], astronomical data analysis [18].

In this paper, we analyze the Barankin bound (BB) [19] for multiple change-point estimation in the context of an independent vector sequence. The Barankin bound is the greatest lower bound for any unbiased estimator. Moreover, in contrast to the CRB, its computation is not limited by the discrete nature of the parameter and the regularity assumptions on the likelihood of the observations [13], [20]. However, the BB requires the use of parameters called test points. These test points choice is left to the user, and, in order to obtain the best (*i.e.*, the tightest) bound, a nonlinear maximization over these test points has to be performed. This explains why this bound is so much less used and known than the CRB, nevertheless, the BB is often a practical bound for realistic scenarios, see *e.g.* [21].

To the best of our knowledge, minimal bounds other than the CRB have been proposed in the context of change-point estimation only in the foundational communication of Ferrari and Tourneret [22]. A simplified and practical version of the BB (*i.e.* one test point per parameter), the so-called Hammersley-Chapman-Robbins (HCR) bound, [20], [23], is studied in that paper. As in the previous works on the CRB, only one change point is considered.

In this paper we extend the results presented in [22] to the case of multiple change points. We consider the multi-parameter HCR bound and we show that the so-called Barankin information matrix (BIM), which has to be inverted, has an interesting structure (*viz.*, a block diagonal matrix structure). We show that the estimation of one change point is corrupted by its neighboring change points and we give the details of the computation for the two change-point case. This case facilitates the derivation of a closed-form expression for the inverse of the BIM. Note that it is possible to find tighter bounds by using more test-points per parameter, however, such approach does not allow for obtaining closed-form expressions of the BIM and its inverse as derived here. We also discuss on the existence of the supremum of the finite set formed by all possible BB solutions and, following ideas from [24] and from convex optimization, we compute a suitable minimal-upper bound to this candidate set with respect to the Loewner cone, the set of semipositive definite matrices. In particular, we show that its computation is given by the matrix associated with the Lowner-John Ellipsoid of the candidate set, which is the minimum-volume hyper-ellipsoid covering the set of hyper-ellipsoids associated to each matrix in the candidate set. We apply the bounds to the case of changes in the parameters of Gaussian and Poisson observations. We finally present

numerical examples for comparing our bound to the performance achieved by the maximum likelihood estimator (MLE).

The notational convention adopted in this paper is as follows: italic indicates a scalar quantity, as in A; lowercase boldface indicates a vector quantity, as in $\mathbf{a}$; uppercase boldface indicates a matrix quantity, as in $\mathbf{A}$. The matrix transpose is indicated by a superscript ^{T} as in $\mathbf{A}^T$. The m^{th}-row and n^{th}-column element of the matrix $\mathbf{A}$ is denoted by $[\mathbf{A}]_{mn}$. The identity matrix of size $N \times N$ is denoted $\mathbf{I}_N$. We define by $\mathbf{1}_{M\times N}$ the matrix such that $[\mathbf{1}]_{mn} = 1$, $\forall m = 1 \dots M$ and $\forall n = 1 \dots N$, and $\mathbf{D}(\mathbf{a})$ is a diagonal matrix formed by the elements of the row vector $\mathbf{a}$. The trace operator is defined as $Tr\{.\}$. The determinant of a matrix is denoted by $|.|$ and cardinality when applying to a set. $\mathbb{S}^n$ denotes the vector space of symmetric $n \times n$ matrices and the subsets of nonnegative definite matrices and positive definite matrices are denoted by $\mathbb{S}^n_+$ and $\mathbb{S}^n_{++}$, respectively. The notation $\mathbf{A} \succeq \mathbf{B}$ means that for $\mathbf{A}$, $\mathbf{B} \in \mathbb{S}^n$, $\mathbf{A} - \mathbf{B} \in \mathbb{S}^n_+$, also known as *Loewner partial ordering* of symmetric matrices [25], [26]. The absolute value is denoted by $abs(.)$. The indicator function of a set S is denoted by $I_S(.)$. The expectation operator is denoted by $\mathbb{E}[.]$. The observation and parameter spaces are denoted, respectively, by Ω and Θ.

The remainder of this paper is organized as follows: In Section II, we present the signal model, the assumptions, and we introduce the general structure for Barankin bound for the signal model parameters. The computation and analysis of the Barankin bound for the change-point localization parameters are provided in Section III. In Section IV, we analyze the cases of changes in the parameters of Gaussian and Poisson distributions. To illustrate our results, simulations are presented in Section V. Finally, in Section VI we conclude this work.

II. Problem Formulation

A. Observation model

We consider the general case of N independent vector observations $\mathbf{X} = [\mathbf{x}_1,\ \mathbf{x}_2,\ \dots,\ \mathbf{x}_N] \in \mathbb{R}^{M\times N}$, which can be obtained, for example, by a multiple sensor system and are modeled as follows:

$$\begin{cases} \mathbf{x}_i \sim p_1(\mathbf{x}_i;\boldsymbol{\eta}_1) \ \text{ for } i = 1,\dots,t_1, \\ \mathbf{x}_i \sim p_2(\mathbf{x}_i;\boldsymbol{\eta}_2) \ \text{ for } i = t_1+1,\dots,t_2, \\ \vdots \\ \mathbf{x}_i \sim p_{q+1}(\mathbf{x}_i;\boldsymbol{\eta}_{q+1}) \ \text{ for } i = t_q+1,\dots,N, \end{cases} \tag{1}$$

where M is the size of the sample vector (*e.g.*, the number of sensors), q is the number of change-points, and p_j is a probability density function (or mass function for discrete random variables) with

parameters $\boldsymbol{\eta}_j \in \mathbb{R}^L$. In other words, $\mathbf{x}_i \sim p_j\left(\mathbf{x}_i;\boldsymbol{\eta}_j\right)$ for $i = t_{j-1}+1,\ \ldots,\ t_j$, with $j = 1,\ \ldots,\ q+1$,where we define $t_0 = 0$ and $t_{q+1} = N$. Note that if $M = 1$, the problem is reduced to the estimation of changes in a time series. We assume that all probability density functions p_j belong to a common distribution. The unknown parameters of interest are the change-point locations $\{t_1, t_2, \ldots, t_q\}$ with $\{t_k \in \mathbb{N} - \{0\}, \quad k = 1, \ldots,\ q\}$, $1 < t_1 < t_2 < \cdots < t_q < N$, and $q < N-2$. The observations between two consecutive change points are assumed to be stationary. Consequently, the $q \times 1$ vector of unknown true parameters for this model is $\mathbf{t} = [t_1,\ t_2,\ \ldots,\ t_q]^T$.

The observation model (1) is useful in signal processing; several examples were already mentioned in the Introduction. Note that, since we focus on the change-point estimation, we assume that the parameters $\boldsymbol{\eta}_j$ are known. The resulting bound will still be useful if these parameters are unknown, but overly optimistic. Moreover, the complexity of the bound derivation increases for unknown $\boldsymbol{\eta}_j$ and therefore we do not consider this case in this work.

B. Barankin Bound

The P-order BB of a vector $\boldsymbol{\theta}_0 \in \mathbb{R}^q$, denoted by $\mathbf{BB}_P(\boldsymbol{\theta}_0, \mathbf{H}_{q\times P})$, is given as follows (see [27]–[30] for more details):

$$\mathbf{Cov}(\widehat{\boldsymbol{\theta}}) \succeq \mathbf{BB}_P(\boldsymbol{\theta}_0, \mathbf{H}_{q\times P}) = \mathbf{H}_{q\times P}(\boldsymbol{\Phi} - \mathbf{1}_{P\times P})^{-1}\mathbf{H}_{q\times P}^T, \tag{2}$$

where $\mathbf{Cov}(\widehat{\boldsymbol{\theta}})$ is the covariance matrix of an unbiased estimator $\widehat{\boldsymbol{\theta}}$ of the parameter vector $\boldsymbol{\theta}_0$. The matrix $\mathbf{H} = [\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0,\ \ldots,\ \boldsymbol{\theta}_P - \boldsymbol{\theta}_0]$ is a function of the set $\{\boldsymbol{\theta}_1,\ \ldots,\ \boldsymbol{\theta}_P\}$, so-called "test points", left to the user's choice. We define $\mathbf{h}_i = \boldsymbol{\theta}_i - \boldsymbol{\theta}_0$ such that the matrix $\mathbf{H} \in \mathbb{R}^{q\times P}$ becomes $\mathbf{H} = [\mathbf{h}_1,\ \ldots,\ \mathbf{h}_P]$. Moreover, note that $\boldsymbol{\theta}_0 + \mathbf{h}_j \in \Theta$. In the following, for simplicity, we use the term "test point" for the vectors $\mathbf{h}_i$. Finally, $\boldsymbol{\Phi}$ is a $\mathbb{R}^{P\times P}$ matrix whose elements $[\boldsymbol{\Phi}]_{kl}$ are given by:

$$[\boldsymbol{\Phi}]_{kl} = \mathbb{E}[L(\mathbf{X}, \boldsymbol{\theta}_0, \mathbf{h}_k)L(\mathbf{X}, \boldsymbol{\theta}_0, \mathbf{h}_l)], \tag{3}$$

where $L(\mathbf{X}, \boldsymbol{\theta}_0, \mathbf{h}_j)$ is defined by,

$$L(\mathbf{X}, \boldsymbol{\theta}_0, \mathbf{h}_j) = \frac{p(\mathbf{X}; \boldsymbol{\theta}_0 + \mathbf{h}_j)}{p(\mathbf{X}; \boldsymbol{\theta}_0)}, \tag{4}$$

where $p(\mathbf{X}; \boldsymbol{\varphi})$ is the likelihood of the observations with parameter vector $\boldsymbol{\varphi}$. Note that the matrix $\boldsymbol{\Phi} - \mathbf{1}_{P\times P}$ is sometimes referred to as the Barankin information matrix (BIM) [31].

As already stated, test points choice is left to the user, since any set of test points in $\mathbf{BB}_P(\boldsymbol{\theta}_0)$ satisfies the inequality (2). Thus, the tightest BB, denoted by $\mathbf{BB}(\boldsymbol{\theta}_0)$, is given as follows:

$$\mathbf{BB}(\boldsymbol{\theta}_0) = \lim_{P \longrightarrow |\Theta|} \sup_{\{\mathbf{h}_1, \ldots, \mathbf{h}_P\}} \mathbf{BB}_P(\boldsymbol{\theta}_0, \mathbf{H}_{q\times P}) \succeq \mathbf{CRB}(\boldsymbol{\theta}_0), \tag{5}$$

where $|\Theta|$ is the cardinality of the set Θ formed by all possible parameter values, and $\mathbf{CRB}(\boldsymbol{\theta}_0)$ is the CRB of $\boldsymbol{\theta}_0$, which, assuming that it exists, is smaller than the $\mathbf{BB}(\boldsymbol{\theta}_0)$ in the Loewner ordering sense. The computation of $\mathbf{BB}(\boldsymbol{\theta}_0)$ is costly, since the limit on P usually implies that a large, possibly infinite, number of test points needs to be considered, a nonlinear maximization over the test points has to be performed, and the inverse of the BIM has to be computed.

Concerning the BB for the parameter vector $\boldsymbol{\theta}_0 = \mathbf{t}$, $|\Theta|$ depends on the number of samples N and change points q as follows:

$$|\Theta| = \sum_{t_1=1}^{N-q} \sum_{t_2=t_1+1}^{N-q+1} \cdots \sum_{t_{q-1}=t_{q-2}+1}^{N-1} (N - t_{q-1} - 1) = \binom{N-1}{q}. \tag{6}$$

Note that $|\Theta| \to \infty$, as $(N-q) \to \infty$ and for N finite then $|\Theta|$ is finite. In practice, the number of test points and the particular structure of matrix $\mathbf{H}$ is usually chosen based on the analytical and computational complexity associated to it, which lead to approximated versions of the BB. In the latter case it would be useful to have some knowledge on how different Barankin bound approximations compare among each other *w.r.t. Loewner partial ordering.* In the following proposition we provide with a general guideline for this purpose:

Lemma 1: Let $\mathbf{A} \in \mathbb{S}_{++}^q$, $\mathbf{B} \in \mathbb{S}_{+}^q$ with $rank\,(\mathbf{B}) = m < q$, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m > 0$ and $\lambda_{m+1} = \cdots = \lambda_q = 0$ the roots of the characteristic equation $|\mathbf{B} - \lambda\mathbf{A}| = 0$. If $\lambda_1 \leq 1$, then $\mathbf{A} \succ \mathbf{B}$ otherwise $\mathbf{A}$ and $\mathbf{B}$ are not mutually comparable.

Proof: See Appendix A ■

If $rank\,(\mathbf{H}_{q\times P}) = q$ then $\mathbf{BB}_P(\boldsymbol{\theta}_0, \mathbf{H}_{q\times P}) \in \mathbb{S}_{++}^q$ since $(\boldsymbol{\Phi} - 1)^{-1} \in \mathbb{S}_{++}^q$ by construction, and if $rank\,(\mathbf{H}_{q\times P'}) < q$ then $\mathbf{BB}_P(\boldsymbol{\theta}_0, \mathbf{H}_{q\times P'}) \in \mathbb{S}_{+}^q$. The Lemma can now be used with $\mathbf{A} = \mathbf{BB}_P(\boldsymbol{\theta}_0, \mathbf{H}_{q\times P})$ and $\mathbf{B} = \mathbf{BB}_{P'}(\boldsymbol{\theta}_0, \widetilde{\mathbf{H}}_{q\times P'})$ provided $rank\,(\mathbf{H}_{q\times P}) = q > rank\left(\widetilde{\mathbf{H}}_{q\times P'}\right)$. Note that $rank\left(\widetilde{\mathbf{H}}_{q\times P'}\right) < q$ implies that the number of test-points $P' < q$, therefore, a matrix bound $\mathbf{BB}_{P'}(\boldsymbol{\theta}_0, \widetilde{\mathbf{H}}_{q\times P'})$ cannot be larger, *w.r.t. Loewner partial ordering,* than any matrix bound given by a test-point matrix $\mathbf{H}_{q\times P}$ consisting of $P = q$ independent test-point vectors. Consequently, in the following we will use an approximate version of the BB that allows us to derive efficiently computed closed-form expressions for the BIM and its inverse in the context of our multiple change-point estimation problem. In particular, we will compute the multi-parameter HCR bound [27] with the classical assumption of one test point per parameter $(P = q)$, i.e., $\mathbf{h}_j = [0,\ \ldots,\ \alpha_j,\ \ldots,\ 0]^T$. Then, $\mathbf{H}$ is a diagonal matrix given by,

$$\mathbf{H} = [\mathbf{h}_1,\ \ldots,\ \mathbf{h}_q] = \mathbf{D}\,(\boldsymbol{\alpha}) \tag{7}$$

where the vector $\boldsymbol{\alpha} = [\alpha_1,\ \ldots,\ \alpha_q]^T$ corresponds to the set of test points associated to the parameters

$\mathbf{t} = [t_1,\ t_2,\ \ldots,\ t_q]^T$. Note that $\alpha_j \neq 0$ is defined such that $t_j + \alpha_j$ ranges over all possible values of t_j, for $j = 1, \ldots, q$. Thus, $\alpha_j \in \{\mathbb{Z} \cap [t_{j-1} - t_j + 1,\ t_{j+1} - t_j - 1] - \{0\}\}$. Let $S \subset \mathbb{Z}^q$ be a set formed by all possible values of $\boldsymbol{\alpha}$. The set S is finite given that t_{q+1} is finite.

The matrix, $\boldsymbol{\Phi} - \mathbf{1}_{q\times q}$, corresponds to the BIM for change-point locations $\mathbf{t}$, denoted here by $\mathbf{BIM_t}$. The approximated BB, $\mathbf{BB}_{\mathbf{t},\,q}$, is then obtained from

$$\mathbf{BB}_{\mathbf{t},\,q} = \sup_{[\mathbf{h}_1,\ldots,\mathbf{h}_q]} \mathbf{BB}_q(\mathbf{t}, \mathbf{H}_{q\times q}) = \sup_{\boldsymbol{\alpha}\,\in\, S} \mathbf{D}\left(\boldsymbol{\alpha}\right)\ \mathbf{BIM}_{\mathbf{t}}^{-1}\mathbf{D}\left(\boldsymbol{\alpha}\right)^T. \tag{8}$$

By construction, the finite set $C := \{\mathbf{BB}_q(\mathbf{t}, \mathbf{D}\left(\boldsymbol{\alpha}\right)),\ \boldsymbol{\alpha} \in S\}$ is a subset of the partially ordered set $(\mathbb{S}^q, \preceq)$ with partial order "$\preceq$" given by the *Loewner ordering* [25], [26]. This partial order is not a lattice ordering, *i.e.*, each finite subset of $\mathbb{S}^q$ may not be closed under least-upper (infimum) and greatest-lower bounds (supremum) [26]. In other words, the notion of a unique supremum or an infimum of C might not exist with respect to the Loewner ordering. The supremum does not exist if there is no upper bound to the set, or if the set of upper bounds does not have a least element. If the supremum exists, it does not need to be defined in the set, but if it belongs to it, then it is the greatest element[1] in the set. Note that a set with respect to the partially order set $(\mathbb{S}^q, \succeq)$ may have several maximal[2] and minimal elements without having a greatest and least element in the set, respectively. If the set has a greatest or least element then it is the unique maximal or minimal element and therefore it is the supremum or infimum of the set. Here, we will approach the computation of the supremum by computing a suitable minimal element of the set of upper bounds of C, namely, a minimal-upper bound $\mathbf{B}_q \in \mathbb{S}^q_{++}$ such that $\mathbf{B}_q \succeq C$ and which is minimal in the sense that there is not smaller matrix $\mathbf{B}'_q \preceq \mathbf{B}_q$ such that $\mathbf{B}'_q \succeq C$. From Eq. (2), $\mathbf{Cov}(\widehat{\boldsymbol{\theta}})$ belongs to set of upper bounds of C therefore if the set of upper bounds has a unique minimal element, i.e., a least element, then $\mathbf{Cov}(\widehat{\boldsymbol{\theta}}) \succeq \mathbf{B}_q$. However, if the set of upper bounds has several minimal elements then in general we can expect that $\mathbf{Cov}(\widehat{\boldsymbol{\theta}}) \succeq \mathbf{B}_q$, or that $\mathbf{Cov}(\widehat{\boldsymbol{\theta}})$ and $\mathbf{B}_q$ are not mutually comparable.

Having a closed form for $\mathbf{BIM}_{\mathbf{t}}^{-1}$ makes the task of computing $\mathbf{B}_q$ much less computationally demanding than having to invert $\mathbf{BIM_t}$ for every $\boldsymbol{\alpha} \in S$. In the following section, we will first derive the elements of $\mathbf{BIM_t}$ and obtain closed-form expressions for $\mathbf{BIM}_{\mathbf{t}}^{-1}$. Then, we will introduce the approach for computing the minimal-upper bound $\mathbf{B}_q$.

[1] $\mathbf{B}_i \in C$ is the greatest element of C *w.r.t.* $(\mathbb{S}^q, \preceq)$ if $\mathbf{B}_i \succeq \mathbf{Y}$ for all $\mathbf{Y} \in C$. If the greatest element exists it is an upper-bound of C contained in it. The least element of C is defined similarly considering $\mathbf{B}_i \preceq \mathbf{Y}$.

[2] $\mathbf{B}_i \in C$ is a maximal element of C *w.r.t.* $(\mathbb{S}^q, \preceq)$ if there is not $\mathbf{Y} \in C$ such that $\mathbf{Y} \succeq \mathbf{B}_i$ and is a minimal element if there is not $\mathbf{Y} \in C$ such that $\mathbf{B}_i \succeq \mathbf{Y}$.

III. BARANKIN BOUND TYPE FOR MULTIPLE CHANGE-POINT ESTIMATION

To compute the BB for the change point localization parameters, we first need to compute $\mathbf{BIM_t}$, which depends on the matrix $\mathbf{\Phi}$. From Equations (3) and (4), the elements of $[\mathbf{\Phi}]_{kl}$, for $k,\, l = 1, \ldots, q$ are given by

$$[\mathbf{\Phi}]_{kl} = \int_{\Omega} \frac{p\left(\mathbf{X}; \mathbf{t} + \mathbf{h}_k\right) p\left(\mathbf{X}; \mathbf{t} + \mathbf{h}_l\right)}{p\left(\mathbf{X}; \mathbf{t}\right)} d\mathbf{X}, \tag{9}$$

where $p\left(\mathbf{X}; \mathbf{t}\right)$ is given by

$$p\left(\mathbf{X}; \mathbf{t}\right) = \prod_{i=1}^{t_1} p_1(\mathbf{x}_i; \boldsymbol{\eta}_1) \cdots \prod_{i=t_{k-1}+1}^{t_k} p_k(\mathbf{x}_i; \boldsymbol{\eta}_k) \cdots \prod_{i=t_q+1}^{N} p_{q+1}(\mathbf{x}_i; \boldsymbol{\eta}_{q+1}), \tag{10}$$

and $p\left(\mathbf{X}; \mathbf{t} + \mathbf{h}_k\right)$ is given by

$$p\left(\mathbf{X}; \mathbf{t} + \mathbf{h}_k\right) = \prod_{i=1}^{t_1} p_1(\mathbf{x}_i; \boldsymbol{\eta}_1) \cdots \prod_{i=t_{k-1}+1}^{t_k+\alpha_k} p_k(\mathbf{x}_i; \boldsymbol{\eta}_k) \cdots \prod_{i=t_q+1}^{N} p_{q+1}(\mathbf{x}_i; \boldsymbol{\eta}_{q+1}), \tag{11}$$

and where $p\left(\mathbf{X}; \mathbf{t} + \mathbf{h}_l\right)$ is same as Equation (11) ($k = l$).

In order to study and to simplify $\mathbf{\Phi}$, we will analyze its diagonal and non-diagonal elements separately.

A. Diagonal elements of $\mathbf{\Phi}$

Replacing $k = l$ in (9) and using (11), we obtain the following expression:

$$[\mathbf{\Phi}]_{kk} = \int_{\Omega} \frac{\prod_{i=1}^{t_1} p_1^2(\mathbf{x}_i; \boldsymbol{\eta}_1) \cdots \prod_{i=t_{k-1}+1}^{t_k+\alpha_k} p_k^2(\mathbf{x}_i; \boldsymbol{\eta}_k) \cdots}{\prod_{i=1}^{t_1} p_1(\mathbf{x}_i; \boldsymbol{\eta}_1) \cdots \prod_{i=t_q+1}^{N} p_{q+1}(\mathbf{x}_i; \boldsymbol{\eta}_{q+1})} \cdots \prod_{i=t_q+1}^{N} p_{q+1}^2(\mathbf{x}_i; \boldsymbol{\eta}_{q+1}) d\mathbf{X}. \tag{12}$$

This equation can be further simplified by considering the the cases $\alpha_k > 0$ and $\alpha_k < 0$, obtaining the following expression (see Appendix B for details on its derivation):

$$[\mathbf{\Phi}]_{kk} = \begin{cases} \left(\int_{\Omega} \frac{p_k^2(\mathbf{x}; \boldsymbol{\eta}_k)}{p_{k+1}(\mathbf{x}; \boldsymbol{\eta}_{k+1})} d\mathbf{x} \right)^{\alpha_k}, & \text{if } \alpha_k > 0, \\ \left(\int_{\Omega} \frac{p_{k+1}^2(\mathbf{x}; \boldsymbol{\eta}_{k+1})}{p_k(\mathbf{x}; \boldsymbol{\eta}_k)} d\mathbf{x} \right)^{-\alpha_k}, & \text{if } \alpha_k < 0. \end{cases} \tag{13}$$

B. Non-diagonal elements of $\mathbf{\Phi}$

The computation of the off-diagonal elements of $\mathbf{\Phi}$ can be simplified by using the fact that the matrix $\mathbf{\Phi}$ is symmetric; therefore, we can focus on either the upper or lower triangular part of $\mathbf{\Phi}$. In our derivations below we consider the upper triangular part, *i.e.* $k < l$, then by using (9) and (11), we obtain the following expression for the elements of $\mathbf{\Phi}$:

$$\begin{aligned} [\mathbf{\Phi}]_{kl} \;=\; & \int_{\Omega} \frac{\prod_{i=1}^{t_1} p_1(\mathbf{x}_i; \boldsymbol{\eta}_1) \cdots \prod_{i=t_{k-1}+1}^{t_k+\alpha_k} p_k(\mathbf{x}_i; \boldsymbol{\eta}_k)}{\prod_{i=1}^{t_1} p_1(\mathbf{x}_i; \boldsymbol{\eta}_1) \cdots \cdots \prod_{i=t_q+1}^{N} p_{q+1}(\mathbf{x}_i; \boldsymbol{\eta}_{q+1})} \prod_{i=t_k+\alpha_k+1}^{t_{k+1}} p_{k+1}(\mathbf{x}_i; \boldsymbol{\eta}_{k+1}) \cdots \prod_{i=t_q+1}^{N} p_{q+1}(\mathbf{x}_i; \boldsymbol{\eta}_{q+1}) \\ & \times \prod_{i=1}^{t_1} p_1(\mathbf{x}_i; \boldsymbol{\eta}_1) \cdots \prod_{i=t_{l-1}+1}^{t_l+\alpha_l} p_l(\mathbf{x}_i; \boldsymbol{\eta}_l) \cdots \prod_{i=t_q+1}^{N} p_{q+1}(\mathbf{x}_i; \boldsymbol{\eta}_{q+1}) d\mathbf{X}. \end{aligned} \tag{14}$$

Following the same idea as for the diagonal elements, $[\mathbf{\Phi}]_{kl}$ can be simplified by analyzing the four possible combinations of test-point ranges, namely,

$$\begin{cases} \text{Case 1: } \alpha_k > 0 \text{ and } \alpha_l > 0, \\ \text{Case 2: } \alpha_k < 0 \text{ and } \alpha_l < 0, \\ \text{Case 3: } \alpha_k < 0 \text{ and } \alpha_l > 0, \\ \text{Case 4: } \alpha_k > 0 \text{ and } \alpha_l < 0. \end{cases} \tag{15}$$

For the last case, *i.e.* $\alpha_k > 0$ and $\alpha_l < 0$, two subcases have to be analyzed: (i) $t_k + \alpha_k < t_l + \alpha_l$ and (ii) $t_k + \alpha_k > t_l + \alpha_l$. These two cases correspond to non-overlapping and overlapping test points, respectively. Note that since $k < l$, $t_k < t_l$ and since $\alpha_j \in \{\mathbb{Z} \cap [t_{j-1} - t_j + 1, t_{j+1} - t_j - 1] - \{0\}\}$, the case $t_k + \alpha_k > t_l + \alpha_l$ which corresponds to an overlapping between two test points, can appear only when $l = k + 1$, or, in other words, when we are analyzing two neighboring change points. Then, for Cases 1-3 and subcase (i), Equation (14) becomes (see Appendix C)

$$[\mathbf{\Phi}]_{kl} = 1, \text{ for } l > k \tag{16}$$

and for subcase (ii), keeping in mind that $\alpha_k > 0$ and $\alpha_{k+1} < 0$, Equation (14) becomes

$$[\mathbf{\Phi}]_{kl} = \begin{cases} \left(\int_\Omega \frac{p_k(\mathbf{x};\boldsymbol{\eta}_k) p_{k+2}(\mathbf{x};\boldsymbol{\eta}_{k+2})}{p_{k+1}(\mathbf{x};\boldsymbol{\eta}_{k+1})} d\mathbf{x} \right)^{\beta_k}, & \text{for } l = k+1, \\ 1, & \text{for } l > k+1, \end{cases} \tag{17}$$

where $\beta_k = (t_k + \alpha_k) - (t_{k+1} + \alpha_{k+1})$.

Remark: This last result is fundamental because it proves the natural intuition that the estimation of q change points is not equivalent to q times the estimation of one change point. In other words, it means that the estimation of one change point is perturbed by its two neighbors. We now summarize the previous results.

C. Barankin information matrix $\mathbf{\Phi} - \mathbf{1}_{q \times q}$

Using Equations (13), (16), and (17), it is clear that $\mathbf{BIM_t}$ has at least a tri-diagonal structure:

$$\mathbf{BIM_t} = \begin{bmatrix} A_1 & B_1 & 0 & \cdots & 0 \\ B_1 & A_2 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & A_{q-1} & B_{q-1} \\ 0 & \cdots & 0 & B_{q-1} & A_q \end{bmatrix}, \tag{18}$$

where

$$\begin{aligned} A_k &= [\mathbf{\Phi}]_{kk} - 1, \quad \text{for } k = 1, \ldots, q \\ &= \begin{cases} \left(\int_\Omega \frac{p_k^2(\mathbf{x};\boldsymbol{\eta}_k)}{p_{k+1}(\mathbf{x};\boldsymbol{\eta}_{k+1})} d\mathbf{x} \right)^{\alpha_k} - 1 & \text{if } \alpha_k > 0, \\ \left(\int_\Omega \frac{p_{k+1}^2(\mathbf{x};\boldsymbol{\eta}_{k+1})}{p_k(\mathbf{x};\boldsymbol{\eta}_k)} d\mathbf{x} \right)^{-\alpha_k} - 1 & \text{if } \alpha_k < 0, \end{cases} \end{aligned} \tag{19}$$

and

$$\begin{aligned} B_k &= [\mathbf{\Phi}]_{kk+1} - 1, \quad \text{for } k = 1, \ldots, q-1 \\ &= \begin{cases} 0, & \text{if } \beta_k < 0, \\ \left(\int_\Omega \frac{p_k(\mathbf{x};\boldsymbol{\eta}_k) p_{k+2}(\mathbf{x};\boldsymbol{\eta}_{k+2})}{p_{k+1}(\mathbf{x};\boldsymbol{\eta}_{k+1})} d\mathbf{x} \right)^{\beta_k} - 1, & \text{if } \beta_k > 0. \end{cases} \end{aligned} \tag{20}$$

In the case of one change-point estimation, $\mathbf{BIM_t}$ is reduced to a scalar A_1, and we re-obtain by replacing $\alpha_1 = \alpha$ the result proposed by Ferrari and Tourneret (see Equations (5) and (6) in [22]):

$$A_1 = \begin{cases} \left(\int_\Omega \frac{p_1^2(\mathbf{x};\boldsymbol{\eta}_1)}{p_2(\mathbf{x};\boldsymbol{\eta}_2)} d\mathbf{x} \right)^{\alpha} - 1 & \text{if } \alpha > 0, \\ \left(\int_\Omega \frac{p_2^2(\mathbf{x};\boldsymbol{\eta}_2)}{p_1(\mathbf{x};\boldsymbol{\eta}_1)} d\mathbf{x} \right)^{-\alpha} - 1 & \text{if } \alpha < 0. \end{cases} \tag{21}$$

Note also that the diagonal elements of $\mathbf{BIM_t}$ can be computed numerically in one step (*i.e.*, $\forall \alpha_k \gtrless 0$) as follows:

$$A_k = \left(\int_\Omega \left(\frac{p_k(\mathbf{x};\boldsymbol{\eta}_k)}{p_{k+1}(\mathbf{x};\boldsymbol{\eta}_{k+1})} \right)^{\epsilon_k} p_{k+1}(\mathbf{x};\boldsymbol{\eta}_{k+1}) d\mathbf{x} \right)^{abs(\alpha_k)} - 1, \tag{22}$$

where $\epsilon_k = \frac{1}{2}\left(3\frac{\alpha_k}{abs(\alpha_k)} + 1\right)$.

The next step of our analysis is to compute $(\mathbf{BIM_t})^{-1}$. For a given set of test points, it is clear that $t_k + \alpha_k > t_{k+1} + \alpha_{k+1} \Longrightarrow t_{k+1} + \alpha_{k+1} < t_{k+2} + \alpha_{k+2}$, since $\alpha_j \in \{\mathbb{Z} \cap [t_{j-1} - t_j + 1, t_{j+1} - t_j - 1] - \{0\}\}$. In other words, $\forall k$, if $B_k \neq 0$, then $B_{k+1} = B_{k-1} = 0$; therefore, $\mathbf{BIM_t}$ is block diagonal and the maximum size of one block is 2×2. Since the problem is reduced to finding, at worst, the inverse of several 2×2 matrices with the same structure, we will have a straightforward inversion. In this section, we detail the case of two change points, we give the generalization to two neighboring points and use this to derive a closed-form expression for the inverse of $\mathbf{BIM_t}$ and thus $\mathbf{BB}_q(\mathbf{t}, \mathbf{D}(\boldsymbol{\alpha}))$.

1) The case of two change points: In this case we have $q = 2$, $\mathbf{t} = [t_1,\ t_2]^T$, and $\mathbf{BB}_q(\mathbf{t}, \mathbf{D}(\boldsymbol{\alpha}))$ becomes

$$\mathbf{BB}_2(\mathbf{t}, \mathbf{D}(\boldsymbol{\alpha})) = \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix} \begin{bmatrix} A_1 & B_1 \\ B_1 & A_2 \end{bmatrix}^{-1} \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix} \tag{23}$$

with

$$A_1 = \begin{cases} \left(\int_\Omega \frac{p_1^2(\mathbf{x};\boldsymbol{\eta}_1)}{p_2(\mathbf{x};\boldsymbol{\eta}_2)} d\mathbf{x} \right)^{\alpha_1} - 1 & \text{if } \alpha_1 > 0, \\ \left(\int_\Omega \frac{p_2^2(\mathbf{x};\boldsymbol{\eta}_2)}{p_1(\mathbf{x};\boldsymbol{\eta}_1)} d\mathbf{x} \right)^{-\alpha_1} - 1 & \text{if } \alpha_1 < 0, \end{cases}, \quad A_2 = \begin{cases} \left(\int_\Omega \frac{p_2^2(\mathbf{x};\boldsymbol{\eta}_2)}{p_3(\mathbf{x};\boldsymbol{\eta}_3)} d\mathbf{x} \right)^{\alpha_2} - 1 & \text{if } \alpha_2 > 0, \\ \left(\int_\Omega \frac{p_3^2(\mathbf{x};\boldsymbol{\eta}_3)}{p_2(\mathbf{x};\boldsymbol{\eta}_2)} d\mathbf{x} \right)^{-\alpha_2} - 1 & \text{if } \alpha_2 < 0, \end{cases} \tag{24}$$

$$B_1 = \begin{cases} 0 & \text{if } \beta_1 < 0, \\ \left(\int_\Omega \frac{p_1(\mathbf{x};\boldsymbol{\eta}_1) p_3(\mathbf{x};\boldsymbol{\eta}_3)}{p_2(\mathbf{x};\boldsymbol{\eta}_2)} d\mathbf{x} \right)^{\beta_1} - 1, & \text{if } \beta_1 > 0, \end{cases} \tag{25}$$

where $\beta_1 = (t_1 + \alpha_1) - (t_2 + \alpha_2)$.

Consequently, depending on the given set of test points, the following five combinations, corresponding, respectively, to cases: 1, 2, 3, 4 in (15), are possible for $\mathbf{BB}_2(\mathbf{t}, \mathbf{D}(\boldsymbol{\alpha}))$:

$$\left\{ \begin{bmatrix} \frac{\alpha_1^2}{\Delta_{112}^{\alpha_1}-1} & 0 \\ 0 & \frac{\alpha_2^2}{\Delta_{223}^{\alpha_2}-1} \end{bmatrix}, \begin{bmatrix} \frac{\alpha_1^2}{\Delta_{221}^{abs(\alpha_1)}-1} & 0 \\ 0 & \frac{\alpha_2^2}{\Delta_{332}^{abs(\alpha_2)}-1} \end{bmatrix}, \begin{bmatrix} \frac{\alpha_1^2}{\Delta_{221}^{abs(\alpha_1)}-1} & 0 \\ 0 & \frac{\alpha_2^2}{\Delta_{223}^{\alpha_2}-1} \end{bmatrix}, \begin{bmatrix} \frac{\alpha_1^2}{\Delta_{112}^{\alpha_1}-1} & 0 \\ 0 & \frac{\alpha_2^2}{\Delta_{332}^{abs(\alpha_2)}-1} \end{bmatrix}, \right.$$
$$\left. \kappa^{-1} \begin{bmatrix} \alpha_1^2 \left(\Delta_{332}^{abs(\alpha_2)} - 1 \right) & \alpha_1 \alpha_2 \left(1 - \Delta_{132}^{\beta_1} \right) \\ \alpha_1 \alpha_2 \left(1 - \Delta_{132}^{\beta_1} \right) & \alpha_2^2 \left(\Delta_{112}^{\alpha_1} - 1 \right) \end{bmatrix} \right\} \tag{26}$$

where we define $\Delta_{ijk} = \int_\Omega \frac{p_i(\mathbf{x};\boldsymbol{\eta}_i) p_j(\mathbf{x};\boldsymbol{\eta}_j)}{p_k(\mathbf{x};\boldsymbol{\eta}_k)} d\mathbf{x}$,and $\kappa = \left(\Delta_{112}^{\alpha_1} - 1 \right) \left(\Delta_{332}^{abs(\alpha_2)} - 1 \right) - \left(\Delta_{132}^{\beta_1} - 1 \right)^2$.

2) Generalization to q change points: Note that for more change points the process is the same except that the inversion has to be computed, because of the increase of possibilities. However, the matrix to be inverted is block diagonal, with block of size 1×1 or 2×2, as stated in the previous section. In particular, depending on the values of $\boldsymbol{\alpha}$, the elements of $\left[\mathbf{BIM}_\mathbf{t}^{-1}\right]_{kl}$ for $1 < k < q$ and $l = \{k,\, k+1\}$, with $\mathbf{BIM}_\mathbf{t}$, A_k, and B_k given by Equations (18), (19), and (20), respectively, and $\alpha_0 = \alpha_q = 0$ and $B_q = 0$, have the following possible values:

If $t_{k+1} + \alpha_{k+1} < t_k + \alpha_k$, then $\alpha_k > 0$, $B_k \neq 0$ and $B_{k-1} = B_{k+1} = 0$, thus

$$\left[\mathbf{BIM}_\mathbf{t}^{-1}\right]_{kl} = \begin{cases} \frac{A_{k+1}}{A_k A_{k+1} - B_k^2}, & \text{for } l = k, \\ -\frac{B_k}{A_k A_{k+1} - B_k^2}, & \text{for } l = k+1. \end{cases} \tag{27}$$

If $t_k + \alpha_k < t_{k-1} + \alpha_{k-1}$, then $\alpha_k < 0$, $B_{k-1} \neq 0$ and $B_{k-2} = B_k = 0$, thus

$$\left[\mathbf{BIM}_\mathbf{t}^{-1}\right]_{kl} = \begin{cases} \frac{A_{k-1}}{A_k A_{k-1} - B_{k-1}^2}, & \text{for } l = k, \\ 0, & \text{for } l = k+1. \end{cases} \tag{28}$$

If $t_{k-1} + \alpha_{k-1} < t_k + \alpha_k < t_{k+1} + \alpha_{k+1}$, then $B_{k-1} = B_k = 0$, thus

$$\left[\mathbf{BIM}_\mathbf{t}^{-1}\right]_{kl} = \begin{cases} \frac{1}{A_k}, & \text{for } l = k, \\ 0, & \text{for } l = k+1. \end{cases} \tag{29}$$

Therefore, the elements of $\left[\mathbf{BIM}_\mathbf{t}^{-1}\right]_{kl}$ for $k,\, l = 1, \ldots, q$, which is a symmetric matrix, are given by

$$\left[\mathbf{BIM}_\mathbf{t}^{-1}\right]_{kl} = \begin{cases} \frac{A_{k-1} I_{[-\infty,-1]}(\alpha_k) + A_{k+1} I_{[1,\infty]}(\alpha_k)}{A_k (A_{k-1} I_{[-\infty,-1]}(\alpha_k) + A_{k+1} I_{[1,\infty]}(\alpha_k)) - \left(B_{k-1}^2 + B_k^2 \right)}, & \text{for } l = k, \\ \frac{-B_k}{A_k A_{k+1} - B_k^2}, & \text{for } l = k+1, \\ 0, & \text{for } l > k+1. \end{cases} \tag{30}$$

Since the matrix $\mathbf{BB}_q(\mathbf{t},\mathbf{D}(\boldsymbol{\alpha}))=\mathbf{D}(\boldsymbol{\alpha})\,\mathbf{BIM}_{\mathbf{t}}^{-1}\mathbf{D}(\boldsymbol{\alpha})^T$, then, $[\mathbf{BB}_q(\mathbf{t},\mathbf{D}(\boldsymbol{\alpha}))]_{kl}$ for $k,\,l=1,\ldots,q$ is given as follows:

$$[\mathbf{BB}_q(\mathbf{t},\mathbf{D}(\boldsymbol{\alpha}))]_{kl}=\begin{cases}\frac{\alpha_k^2(A_{k-1}I_{[-\infty,-1]}(\alpha_k)+A_{k+1}I_{[1,\infty]}(\alpha_k))}{A_k(A_{k-1}I_{[-\infty,-1]}(\alpha_k)+A_{k+1}I_{[1,\infty]}(\alpha_k))-\left(B_{k-1}^2+B_k^2\right)}, & \text{for } l=k,\\ -\frac{\alpha_k\alpha_{k+1}B_k}{A_kA_{k+1}-B_k^2}, & \text{for } l=k+1,\\ 0, & \text{for } l>k+1,\end{cases} \tag{31}$$

where $[\mathbf{BB}_q(\mathbf{t},\mathbf{D}(\boldsymbol{\alpha}))]_{kl}=[\mathbf{BB}_q(\mathbf{t},\mathbf{D}(\boldsymbol{\alpha}))]_{lk}$. If for a given set of test points there is no overlap with the neighboring change-point t_{k-1} and t_{k+1}, then, $B_{k-1}=B_k=0$ in (31) and we obtain the particular result $[\mathbf{BB}_q(\mathbf{t},\mathbf{D}(\boldsymbol{\alpha}))]_{kk}=\alpha_k^2/A_k$ and $[\mathbf{BB}_q(\mathbf{t},\mathbf{D}(\boldsymbol{\alpha}))]_{kk+1}=[\mathbf{BB}_q(\mathbf{t},\mathbf{D}(\boldsymbol{\alpha}))]_{k+1k}=0$. This is equivalent to the bound obtained using the same set of test points and assuming one change-point located in the time interval between t_{k-1} and t_{k+1} with total numbers of time-samples $N=t_{k+1}-t_{k-1}$.

D. Computation of the supremum

To obtain the tightest bound from the finite set $C:=\{\mathbf{BB}_q(\mathbf{t},\mathbf{D}(\boldsymbol{\alpha})),\ \boldsymbol{\alpha}\in S\}\subset\mathbb{S}_{++}^q$ we need to compute the supremum of C with respect to the partially order set $(\mathbb{S}^q,\succeq)$. The partial order is given by the Loewner ordering which is defined via the cone of positive semidefinite matrices [25], [26]. In general, this problem is indeed very complex since it requires to look for $\boldsymbol{\alpha}^*\in S$ such that $\mathbf{BB}_q(\mathbf{t},\mathbf{D}(\boldsymbol{\alpha}^*))\succeq\mathbf{BB}_q(\mathbf{t},\mathbf{D}(\boldsymbol{\alpha}))$ for all $\boldsymbol{\alpha}\in S$. To the best of our knowledge, no formal approach for solving this problem has been proposed in the technical literature of minimal bounds. For example, in [28], [32] the choice of the test point $\boldsymbol{\alpha}$ has been guided by some physical considerations of the model being studied. Also, from an optimal design context [25], an approximation for solving this problem is to compute the matrix in C with the largest trace, $\mathbf{BB}_{\mathrm{tr}}$. However, the fact that $Tr\{\mathbf{BB}_{\mathrm{tr}}\}>Tr\{\mathbf{B}_i\}$ for $\mathbf{B}_i\in C$, does not imply that $\mathbf{BB}_{\mathrm{tr}}\succ\mathbf{B}_i$, only the converse statement is valid. In fact, only if C has a greatest element, i.e., the supremum of the set, then it is given by the matrix in C with the largest trace Let $\mathbf{B}_j=\sup C$, with $\mathbf{B}_j\in C$, then by definition $\mathbf{B}_j\succeq\mathbf{B}_i$, for all $\mathbf{B}_i\in C$ with $i\neq j$. Let $\mathbf{G}=\mathbf{B}_j-\mathbf{B}_i$, thus $\mathbf{G}\in\mathbb{S}_+^q$ and $Tr\{\mathbf{G}\}>0$. Hence, $Tr\{\mathbf{B}_j\}>Tr\{\mathbf{B}_i\}$, for all $\mathbf{B}_i\in C$ with $i\neq j$, but as we discussed at the end of Section II, the notion of a unique supremum or an infimum with respect to the Loewner partial ordering in the finite set C might not exist.

Here we address the computation of the supremum by finding a minimal-upper bound $\mathbf{B}_q\in\mathbb{S}_{++}^q$ to the set C such that, $\mathbf{B}_q\succeq C$ and which is minimal in the sense that there is no smaller matrix $\mathbf{B}_q'\preceq\mathbf{B}_q$ such that $\mathbf{B}_q'\succeq C$. In [24], the authors implicitly introduced an algorithm for computing a minimal-upper bound to a finite set of positive definite matrices and redefined this element as supremum of the set.

Before discussing more details about it, we need to introduce the so-called penumbra $P(\mathbf{M})$ of a matrix $\mathbf{M} \in \mathbb{S}^q$ as the set $P(\mathbf{M}) := \{\mathbf{N} \in \mathbb{S}^q : \mathbf{N} \preceq \mathbf{M}\}$ [24], [25] and the following proposition:

Proposition 2: Define $\mathbf{M}$ and $\mathbf{N} \in \mathbb{S}^q$, then $\mathbf{M} \succeq \mathbf{N}$ iff $P(\mathbf{N}) \subseteq P(\mathbf{M})$.

Proof: If $P(\mathbf{N}) \subseteq P(\mathbf{M})$, then $\mathbf{N} \in P(\mathbf{M})$ and then, by definition of penumbra, $\mathbf{M} \succeq \mathbf{N}$. To prove the other implication, we define a matrix $\mathbf{G} \in \mathbb{S}^q$ such as $\mathbf{N} \succeq \mathbf{G}$. Then if $\mathbf{M} \succeq \mathbf{N}$ we have that by the transitivity property of the Loewner order $\mathbf{M} \succeq \mathbf{G}$, namely, $\mathbf{M} \succeq \mathbf{N} \succeq \mathbf{G}$. Therefore, all the matrix elements in $P(\mathbf{N})$ are also in $P(\mathbf{M})$, thus, $P(\mathbf{N}) \subseteq P(\mathbf{M})$. ■

The penumbra $P(\mathbf{M})$ is seen as an inverted cone of vertex $\mathbf{M}$ characterizing all matrices that are smaller than $\mathbf{M}$ [24], [25]. The authors in [24], [25] redefined the supremum of a set of matrices as the matrix associated to the vertex of the minimal penumbra covering the penumbras of all the matrices in the set. The minimal-penumbra vertex is a minimal-upper bound to the set with respect to the partially order set $(\mathbb{S}^q, \succeq)$. In [24], the minimal-penumbra vertex is computed by associating with each matrix $\mathbf{M} \in \mathbb{S}^q$ a ball in the subspace $\mathbb{S}_{\mathbf{A}} = \{\mathbf{A} : Tr\{\mathbf{A}\} = 0\}$, and the authors show that it is determined by the smallest ball enclosing the set of balls associated to each matrix in the set. The latter algorithm is implemented in an approximate manner, by solving instead the problem of finding the smallest enclosing ball of a set of points which correspond to samples from the boundaries of each ball. The success of this method to obtain a minimal-upper bound matrix depends on the samples chosen. For example, in the case of having two balls, it is easy to show that the smallest enclosing ball is tangent to each ball border at the two farthest points from the set of points defined by the intersection of a line passing through each ball center and each ball boundary. Therefore if the sampling procedure does not include this pair of points, then the resulting ball does not enclose completely both balls and, thus, the resulting matrix is not a minimal-upper bound. Moreover, when the dimension is larger than two, a simple analytical computation shows that this algorithm fails to obtain a minimal-upper bound matrix for the set formed by two diagonal matrices no comparable to each other according to Loewner order.

Here, instead, we propose a method for computing a suitable $\mathbf{B}_q$ for any dimension. First, we show that computing $\mathbf{B}_q$ is equivalent to finding the minimum-volume hyper-ellipsoid covering the set of hyper-ellipsoids associated to each matrix in the set C. And second, we show that this problem can be written as a convex objective function with convex constraints which can be solved efficiently using semidefinite programming. An hyper-ellipsoid $\varepsilon \subset \mathbb{R}^q$ with non-empty interior and centered at the origin can be represented by the set $\varepsilon(\mathbf{F}) = \left\{\mathbf{x}^T\mathbf{F}^{-1}\mathbf{x} \leq \mathbf{1}\right\}$, where $\mathbf{F} \in \mathbb{S}^q_{++}$. Suppose $\varepsilon\left(\widetilde{\mathbf{F}}\right)$ is another hyper-ellipsoid similarly represented where $\widetilde{\mathbf{F}} \in \mathbb{S}^q_{++}$. Then, the following statement holds:

Lemma 3: $\mathbf{F} \succeq \widetilde{\mathbf{F}}$ iff $\varepsilon(\mathbf{F}) \supseteq \varepsilon\left(\widetilde{\mathbf{F}}\right)$.

Proof: By the S-procedure [33], we have that $\varepsilon\left(\widetilde{\mathbf{F}}\right) \subseteq \varepsilon(\mathbf{F})$ if and only if there is a $\lambda > 0$ such that

$$\begin{bmatrix} \mathbf{F}^{-1} & 0 \\ 0 & -1 \end{bmatrix} \preceq \lambda \begin{bmatrix} \widetilde{\mathbf{F}}^{-1} & 0 \\ 0 & -1 \end{bmatrix},$$

with equality when $\lambda = 1$, implying the necessary condition $\widetilde{\mathbf{F}} \preceq \mathbf{F}$. ■

Given a finite set of hyper-ellipsoids $C_\varepsilon := \left\{\varepsilon(\mathbf{F}_i) \mid \mathbf{F}_i \in \mathbb{S}^q_{++},\, i = 1, \ldots, R\right\}$, we can always find a unique minimum volume hyper-ellipsoid, $\varepsilon(\mathbf{F}_{\rm jl})$, containing the set C_ε, *i.e*, containing all $\varepsilon(\mathbf{F}_i)$ [33]. Since C_ε is convex, $\varepsilon(\mathbf{F}_{\rm jl})$ is known as the Lowner-John ellipsoid of C_ε [33] and, as we show in the following statement, $\mathbf{F}_{\rm jl}$ is a minimal-upper bound of the set $C_F := \{\, \mathbf{F}_i,\, i = 1, \ldots, R\}$ formed by all the matrices associated to the hyper-ellipsoids in C_ε.

Theorem 4: The matrix $\mathbf{F}_{\rm jl}$, associated to the Lowner-John ellipsoid of the set C_ε, is a minimal-upper bound of the set C_F *w.r.t* to the Loewner partial ordering.

Proof: We will demonstrate this by contradiction. From *Lemma 3* we have that $\mathbf{F}_{\rm jl} \succeq \mathbf{F}_i$, $i = 1, \ldots, R$. Assume that there exists a matrix $\mathbf{F}_{\rm o} \notin C_F$ such that $\mathbf{F}_{\rm jl} \succeq \mathbf{F}_{\rm o} \succeq \mathbf{F}_i$, therefore $\varepsilon(\mathbf{F}_{\rm o}) \supseteq \varepsilon(\mathbf{F}_i)$, for $i = 1, \ldots, R$, and thus $\varepsilon(\mathbf{F}_{\rm o}) \supseteq \bigcup_{i=1}^{P} \varepsilon(\mathbf{F}_i)$. Given that the volume of $\varepsilon(\mathbf{F}_{\rm jl})$ is less than the volume of $\varepsilon(\mathbf{F}_{\rm o})$, since it is the minimum volume hyper-ellipsoid enclosing all $\mathbf{F}_i$, then $|\mathbf{F}_{\rm jl}| \leq |\mathbf{F}_{\rm o}|$, but by construction $\mathbf{F}_{\rm jl} \succeq \mathbf{F}_{\rm o}$, thus $|\mathbf{F}_{\rm jl}| \geq |\mathbf{F}_{\rm o}|$ which is a contradiction. Thus $\mathbf{F}_{\rm o} = \mathbf{F}_{\rm jl}$ and $\mathbf{F}_{\rm jl}$ is a minimal-upper bound of the set C_F. ■

Therefore, computing a minimal-upper bound matrix $\mathbf{B}_q$ of the set $C := \{\mathbf{B}\mathbf{B}_q(\mathbf{t}, \mathbf{D}(\boldsymbol{\alpha})),\ \boldsymbol{\alpha} \in S\} \subset \mathbb{S}^q_{++}$ is equivalent to finding the Lowner-John ellipsoid of the set of hyper-ellipsoids associated to C. This is a particular case of a more general problem of computing the minimum volume hyper-ellipsoid $\varepsilon(\mathbf{B}) = \left\{\mathbf{x}^T\mathbf{B}^{-1}\mathbf{x} + 2\left(\mathbf{B}^{-1/2}\,\mathbf{b}\right)^T\mathbf{x} + \mathbf{b}^T\mathbf{b} \leq \mathbf{1}\right\}$ which covers the union of a set of non centered hyper-ellipsoids parameterized by the quadratic inequalities $\varepsilon_i(\mathbf{B}_i) = \left\{\mathbf{x}^T\mathbf{B}_i^{-1}\mathbf{x} + 2\mathbf{b}_i^T\mathbf{x} + c_i \leq 0\right\}$ for $i = 1, \ldots, m$. This problem can be posed as follows [33]:

$$\max_{\{\mathbf{B},\,\mathbf{b}\}} \left\{\log\left(\det\left(\mathbf{B}^{1/2}\right)\right)\right\} \tag{32}$$

subject to :

$$\begin{aligned} &\tau_1 \geq 0,\ \tau_2 \geq 0, \ldots, \tau_m \geq 0, \\ &\begin{bmatrix} \mathbf{B}^{-1} - \tau_i\mathbf{B}_i^{-1} & \mathbf{B}^{-1/2}\,\mathbf{b} - \tau_i\mathbf{b}_i \\ \left(\mathbf{B}^{-1/2}\,\mathbf{b} - \tau_i\mathbf{b}_i\right)^T & \mathbf{b}^T\mathbf{b} - 1 - \tau_i c_i \end{bmatrix} \preceq 0,\ i = 1, \ldots, m. \end{aligned}$$

The objective function and the set of constrains are convex, so it can be solved efficiently using

semidefinite programming. In particular, we solve this problem using `CVX`, a package for specifying and solving convex programs [34], [35], for $\mathbf{B}_i = \mathbf{BB}_q(\mathbf{t}, \mathbf{D}(\boldsymbol{\alpha}_i))$ for $\boldsymbol{\alpha}_i \in S$, $\mathbf{b}_i = \mathbf{b} = \mathbf{0}$, and $c_i = 1$. Therefore, the minimal-upper bound $\mathbf{B}_q$ of the set C is given by $\mathbf{B}_q = \mathbf{B}_*$, where $\mathbf{B}_*$ is the optimal solution of (32). Using the following statement, we can even reduce the number of constrains in the above problem by considering only the set $C_{\mathrm{m}} \subseteq C$ formed by all the maximal elements of C.

Theorem 5: Define C_{Fm} as the subset of C_F formed by all the maximal elements of C_F. Then, the Lowner-John ellipsoid $\varepsilon(\mathbf{F}_{\mathrm{jl}})$ of C_ε is also the Lowner-John ellipsoid of the set $C_{\varepsilon m}$ formed by the hyper-ellipsoids associated to the matrices in C_{Fm}.

Proof: Since C_{Fm} is formed by all the maximal elements of C_F, then for $\mathbf{F}_i \in C_{Fm}$ and any $\mathbf{F}_j \in C_{Fc} = C_F - C_{Fm}$, we have that $\mathbf{F}_i \succeq \mathbf{F}_j$. From *Lemma 3*, $\varepsilon(\mathbf{F}_i) \supseteq \{\varepsilon(\mathbf{F}_j),\ \text{for all } \mathbf{F}_j \in C_{Fc}\}$, which is true for all $\mathbf{F}_i \in C_{Fm}$, *i.e.*, for all $\varepsilon(\mathbf{F}_i) \in C_{\varepsilon m}$, thus $C_{\varepsilon m} \supseteq \{\varepsilon(\mathbf{F}_j),\ \text{for all } \mathbf{F}_j \in C_{Fc}\}$ and $C_\varepsilon = C_{\varepsilon m} \cup \{\varepsilon(\mathbf{F}_j),\ \text{for all } \mathbf{F}_j \in C_{Fc}\} = C_{\varepsilon m}$. Therefore, $\varepsilon(\mathbf{F}_{\mathrm{jl}})$ is the Lowner-John ellipsoid for the set C_ε and $C_{\varepsilon m}$. ■

Hence, using the above result we decrease the number of constraints in (32) by performing a pre-step which identifies the set C_{m}. Note that if C has a greatest element, it is the unique maximal element of C and therefore it is the supremum of the set and its associated hyper-ellipsoid is the Lowner-John ellipsoid of the set of hyper-ellipsoids associated to C. Therefore, there is not need to solve problem (32). Our algorithm searches and removes from the set of constraints the matrices whose hyper-ellipsoid is fully enclosed by other hyper-ellipsoids. In particular, we evaluate in an iterative manner the membership to C_{m} of all elements in C. We define a membership indicator vector $\mathbf{i}_{C_{\mathrm{m}}}$ where $[\mathbf{i}_{C_{\mathrm{m}}}]_i = I_{C_{\mathrm{m}}}(\mathbf{F}_i)$ and the algorithm begins assuming that all elements belongs to C_{m}, namely, $\mathbf{i}_{C_{\mathrm{m}}} = \mathbf{1}_{R\times 1}$, where $R = |C|$. Then, all the values of the elements of $\mathbf{i}_{C_{\mathrm{m}}}$ are evaluated using the following iterative procedure:

- *Step 0*: Initialize $\mathbf{i}_{C_{\mathrm{m}}} = \mathbf{1}_{R\times 1}$ and set indexes $k = 1$, $l = 1$.
- *Step 1*: Evaluate membership of $\mathbf{F}_k$ to C_{m} (if $k > R$, terminate the algorithm):
$$\text{If } I_{C_{\mathrm{m}}}(\mathbf{F}_k) = \begin{cases} 0, \text{ set } k = k+1 \text{ and restart } \textit{Step 1}, \\ 1, \text{ set } l = l+1 \text{ and go to } \textit{Step 2}. \end{cases}$$
- *Step 2*: Evaluate membership of $\mathbf{F}_l$ to C_{m} (if $l > R$, set $k = k+1$, $l = 1$, and go to *Step 1*):
$$\text{If } I_{C_{\mathrm{m}}}(\mathbf{F}_l) = \begin{cases} 0, \text{ set } l = l+1 \text{ and restart } \textit{Step 2}, \\ 1, \text{ go to } \textit{Step 3}. \end{cases}$$
- *Step 3*: Compare $\mathbf{F}_k$ versus $\mathbf{F}_l$ *w.r.t.* the Loewner ordering:
$$\begin{cases} \text{if } \mathbf{F}_k \succeq \mathbf{F}_l, \text{ set } I_{C_{\mathrm{m}}}(\mathbf{F}_l) = 0,\ l = l+1, \text{ and go to } \textit{Step 2}, \\ \text{if } \mathbf{F}_l \succeq \mathbf{F}_k, \text{ set } I_{C_{\mathrm{m}}}(\mathbf{F}_k) = 0,\ k = k+1,\ l = 1, \text{ and go to } \textit{Step 1}, \\ \text{if } \textit{not comparable}, \text{ set } I_{C_{\mathrm{m}}}(\mathbf{F}_l) = 1,\ l = l+1, \text{ and go to } \textit{Step 2}. \end{cases}$$

Finally, once the algorithm terminates, the set C_{m} will be given by all elements such that $I_{C_{\mathrm{m}}}(\mathbf{F}_i) = 1$. To compare $\mathbf{F}_k$ versus $\mathbf{F}_l$, *w.r.t.* to the Loewner ordering, we apply the determinant test [36] to the matrix, $\mathbf{G} = \mathbf{F}_k - \mathbf{F}_l$. This test evaluates the principal minors of $\mathbf{G}$ and concludes on the matrix definiteness as follows: (i) $\mathbf{G}$ is positive definite, *i.e.*, $\mathbf{F}_k \succ \mathbf{F}_l$, if and only if all its leading principal minors are strictly positive and it is negative definite, *i.e.*, $\mathbf{F}_l \succ \mathbf{F}_k$, if its *k*-th order leading principal minor is < 0 for k odd and > 0 for k even; (ii) $\mathbf{G}$ is positive semidefinite, *i.e.*, $\mathbf{F}_k \succeq \mathbf{F}_l$, if and only if all the principal minors are non-negative and it is negative semidefinite, *i.e.*, $\mathbf{F}_l \succeq \mathbf{F}_k$, if all the *k*-th order principal minors are ≤ 0 for k odd and ≥ 0 for k even; (iii) $\mathbf{G}$ is indefinite, *i.e.*, $\mathbf{F}_k$ and $\mathbf{F}_l$ are not comparable, if none of the previous conditions are satisfied. Since all the matrices in the set C are block diagonal and the maximum size of one block is 2×2, then every matrix $\mathbf{G}$ is a symmetric tridiagonal matrix, which leading principal minors $\{f_{\mathbf{G}}(r),\ r = 1, \ldots, q\}$ can be computed iteratively as follows [37]:

$$f_{\mathbf{G}}(r) = \begin{cases} 1, \text{ for } r = 0, \\ [\mathbf{G}]_{11}, \text{ for } r = 1, \\ [\mathbf{G}]_{r\,r}\ f_{\mathbf{G}}(r-1) - \left([\mathbf{G}]_{r\,r-1}\right)^2 f_{\mathbf{G}}(r-2), \text{ for } 2 < r < q. \end{cases}$$

Note that the determinant of the tridiagonal matrix $\mathbf{G}$ is given by $|\mathbf{G}| = f_{\mathbf{G}}(q)$, and since all the principal minors of $\mathbf{G}$ are also tridiagonal matrices, then their values are computed efficiently using the above expression.

Following the ideas of [24], the issue of having a unique supremum of a set positive definite matrices can be overcome by redefining the supremum as the matrix associated to the Lowner-John Ellipsoid of the set of hyperellipsoids associated to the maximal elements of the set C formed by the P-order BB matrices. This matrix $\mathbf{B}_q$ is unique in the sense that there is not other ellipsoid with minimal volume covering the hyper-ellipsoids associated to the set of maximal element of C. It also has the properties of continuity, namely, it is positive definite. In the following section we will derive the elements of the Barankin information matrix for changes in the parameters of Gaussian and Poisson distributions.

IV. Change in parameters of Gaussian and Poisson distributions

In this section, we apply the proposed bound for two distributions generally encountered in signal processing. We analyze these two cases in a very general way, which means that the results presented here can be applied to a wide variety of estimation problems. Indeed, the parameters involved in the Gaussian distribution (mean and covariance) and in the Poisson distribution are assumed to be a function of the parameters $\boldsymbol{\eta}_j$ which generally represent physical parameters of interest in signal processing.

An example of change of parameters in a Gaussian distribution in the radar context is direction-of-arrival (DOA) estimation. The varying cross-section fluctuations are modeled with a Swerling 0 model [38], where the DOAs are hidden in the mean of the observations, leading for example to the so-called conditional MLE [39]. On the other hand, when the emitted signals are modeled with a Swerling 1-2, the DOAs are hidden in the covariance of the observations, leading, for example, to the so-called unconditional MLE [40]. In the context of particle detection, the Poisson distribution is generally used to model the particle counting process; *i.e.*, the observations and the parameter involved in the Poisson distribution become a function of the DOA [41].

A. Gaussian case

Let us assume that the vector of observations $\mathbf{x}_i \in \mathbb{R}^M$, for $i = 1, \ldots, N$, is modeled as $\mathbf{x}_i = \mathbf{f}(\boldsymbol{\nu}_j) + \mathbf{n}_i$, where, $\mathbf{f}(\cdot)$ is a vector of known functions, $\boldsymbol{\nu}_j \in \mathbb{R}^F$ is a known parameter vector, $\mathbf{n}_i$ is a zero-mean Gaussian random vector with covariance matrix $\mathbf{M}\left(\boldsymbol{\varphi}_j\right)$, with $\mathbf{M}(\cdot)$ is a symmetric positive definite matrix of known functions, and $\boldsymbol{\varphi}_j \in \mathbb{R}^G$ is a known parameter vector. Then $\boldsymbol{\eta}_j = \left[\boldsymbol{\nu}_j^T,\ \boldsymbol{\varphi}_j^T\right]^T \in \mathbb{R}^L$, with $L = F + G$, and $\mathbf{x}_i$ are distributed as $\mathcal{N}\left(\mathbf{f}(\boldsymbol{\nu}_j), \mathbf{M}\left(\boldsymbol{\varphi}_j\right)\right)$. Here we are interested in deriving the elements of the Barankin information matrix for changes in the pdf parameters of $\mathbf{x}_i$, *i.e.*, mean and covariance matrix. First, we analyze the general case of piecewise changes of mean and covariance. Second, we deduce the particular cases i) piecewise changes of mean and constant covariance matrix, *i.e.*, $\mathbf{M}\left(\boldsymbol{\varphi}_j\right) = \mathbf{M}(\boldsymbol{\varphi}) = \boldsymbol{\Sigma}$; ii) piecewise changes of covariance and constant mean vector, *i.e.*, $\mathbf{f}(\boldsymbol{\nu}_j) = \mathbf{f}(\boldsymbol{\nu}) = \boldsymbol{\mu}$. Note that we restrict our analysis to the set of parameter vectors $\{\boldsymbol{\nu}_j\}$ and $\left\{\boldsymbol{\varphi}_j\right\}$ such that the functions in $\mathbf{f}(\boldsymbol{\nu}_j)$ and $\mathbf{M}\left(\boldsymbol{\varphi}_j\right)$ are injective. In other words, a change in the values of $\boldsymbol{\nu}_j$ changes the values of $\mathbf{f}(\boldsymbol{\nu}_j)$, the mean of the distribution of $\mathbf{x}_i$. Similarly a change in the values of $\boldsymbol{\varphi}_j$ implies a change in values of the covariance matrix $\mathbf{M}\left(\boldsymbol{\varphi}_j\right)$. Below, we compute the elements of the Barankin information matrix $\mathbf{BIM_t}$. Then, for each case, respectively, we derive closed-form expressions for the elements $\boldsymbol{\Phi} - \mathbf{1}_{q\times q}$ (see Appendix D for details on their derivation) which are different from zero, namely, we evaluate $[\boldsymbol{\Phi}]_{kk}$ for $\alpha_k > 0$, $\alpha_k < 0$, and $[\boldsymbol{\Phi}]_{kk+1}$for $t_k + \alpha_k > t_{k+1} + \alpha$.

1) Piecewise changes of mean and covariance matrix: For $\alpha_k > 0$, using Equation (13), we have that $[\boldsymbol{\Phi}]_{kk}$ is given by

$$[\boldsymbol{\Phi}]_{kk} = \begin{cases} \left(\frac{\left|\mathbf{M}\left(\boldsymbol{\varphi}_{k+1}\right)\right|^{1/2}\left|\mathbf{M}_k^{-1}\right|^{1/2}}{\left|\mathbf{M}(\boldsymbol{\varphi}_k)\right|}\right)^{\alpha_k} \exp\left\{\frac{\alpha_k}{2}\mathbf{g}_k^T\mathbf{M}_k^{-1}\mathbf{g}_k - \alpha_k\ \mathbf{f}^T(\boldsymbol{\nu}_k)\left(\mathbf{M}\left(\boldsymbol{\varphi}_k\right)\right)^{-1}\mathbf{f}(\boldsymbol{\nu}_k)\right\} \\ \times \exp\left\{\frac{\alpha_k}{2}\mathbf{f}^T(\boldsymbol{\nu}_{k+1})\left(\mathbf{M}\left(\boldsymbol{\varphi}_{k+1}\right)\right)^{-1}\mathbf{f}(\boldsymbol{\nu}_{k+1})\right\},\ \text{for } \mathbf{M}_k \in \mathbb{S}_{++}^M, \\ \infty,\ \text{otherwise,} \end{cases} \tag{33}$$

where $\mathbf{M}_k = \left(2\left(\mathbf{M}\left(\boldsymbol{\varphi}_k\right)\right)^{-1} - \left(\mathbf{M}\left(\boldsymbol{\varphi}_{k+1}\right)\right)^{-1}\right)$ and $\mathbf{g}_k = 2\left(\mathbf{M}\left(\boldsymbol{\varphi}_k\right)\right)^{-1}\mathbf{f}(\boldsymbol{\nu}_k) - \left(\mathbf{M}\left(\boldsymbol{\varphi}_{k+1}\right)\right)^{-1}\mathbf{f}(\boldsymbol{\nu}_{k+1})$.

For $\alpha_k < 0$, using Equation (13), we have that $[\boldsymbol{\Phi}]_{kk}$ is given by

$$[\boldsymbol{\Phi}]_{kk} = \begin{cases} \left(\frac{|\mathbf{M}(\boldsymbol{\varphi}_k)|^{1/2}|\mathbf{M}_{k+1}^{-1}|^{1/2}}{|\mathbf{M}(\boldsymbol{\varphi}_{k+1})|}\right)^{-\alpha_k} \exp\left\{\frac{-\alpha_k}{2}\mathbf{g}_{k+1}^T\mathbf{M}_{k+1}^{-1}\mathbf{g}_{k+1} + \alpha_k\, \mathbf{f}^T(\boldsymbol{\nu}_{k+1})\left(\mathbf{M}\left(\boldsymbol{\varphi}_{k+1}\right)\right)^{-1}\mathbf{f}(\boldsymbol{\nu}_{k+1})\right\} \\ \times \exp\left\{\frac{-\alpha_k}{2}\mathbf{f}^T(\boldsymbol{\nu}_k)\left(\mathbf{M}\left(\boldsymbol{\varphi}_k\right)\right)^{-1}\mathbf{f}(\boldsymbol{\nu}_k)\right\},\ \text{for } \mathbf{M}_{k+1} \in \mathbb{S}_{++}^M, \\ \infty,\ \text{otherwise,} \end{cases} \tag{34}$$

where $\mathbf{M}_{k+1} = 2\left(\mathbf{M}\left(\boldsymbol{\varphi}_{k+1}\right)\right)^{-1} - \left(\mathbf{M}\left(\boldsymbol{\varphi}_k\right)\right)^{-1}$ and $\mathbf{g}_{k+1} = 2\left(\mathbf{M}\left(\boldsymbol{\varphi}_{k+1}\right)\right)^{-1}\mathbf{f}(\boldsymbol{\nu}_{k+1}) - \left(\mathbf{M}\left(\boldsymbol{\varphi}_k\right)\right)^{-1}\mathbf{f}(\boldsymbol{\nu}_k)$.

For $t_k + \alpha_k > t_{k+1} + \alpha_{k+1}$, using Equation (17), we have that $[\boldsymbol{\Phi}]_{kk+1}$is given as follows:

$$[\boldsymbol{\Phi}]_{kk+1} = \begin{cases} \left(\frac{\left|\mathbf{M}(\boldsymbol{\varphi}_{k+1})\right|^{1/2}|\overline{\mathbf{M}}_k^{-1}|^{1/2}}{|\mathbf{M}(\boldsymbol{\varphi}_k)|^{1/2}\left|\mathbf{M}(\boldsymbol{\varphi}_{k+2})\right|^{1/2}}\right)^{\beta_k} \exp\left\{\frac{\beta_k}{2}\overline{\mathbf{g}}_k^T\overline{\mathbf{M}}_k^{-1}\overline{\mathbf{g}}_k^T - \frac{\beta_k}{2}\mathbf{f}^T(\boldsymbol{\nu}_k)\left(\mathbf{M}\left(\boldsymbol{\varphi}_k\right)\right)^{-1}\mathbf{f}(\boldsymbol{\nu}_k)\right\} \\ \times \exp\left\{-\frac{\beta_k}{2}\mathbf{f}^T(\boldsymbol{\nu}_{k+2})\left(\mathbf{M}\left(\boldsymbol{\varphi}_{k+2}\right)\right)^{-1}\mathbf{f}(\boldsymbol{\nu}_{k+2}) + \frac{\beta_k}{2}\mathbf{f}^T(\boldsymbol{\nu}_{k+1})\left(\mathbf{M}\left(\boldsymbol{\varphi}_{k+1}\right)\right)^{-1}\mathbf{f}(\boldsymbol{\nu}_{k+1})\right\}, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{for } \overline{\mathbf{M}}_k^{-1} \in \mathbb{S}_{++}^M, \\ \infty, \text{ otherwise,} \end{cases} \tag{35}$$

where $\overline{\mathbf{M}}_k = \left(\mathbf{M}\left(\boldsymbol{\varphi}_k\right)\right)^{-1} + \left(\mathbf{M}\left(\boldsymbol{\varphi}_{k+2}\right)\right)^{-1} - \left(\mathbf{M}\left(\boldsymbol{\varphi}_{k+1}\right)\right)^{-1}$, and
$\overline{\mathbf{g}}_k = \left(\mathbf{M}\left(\boldsymbol{\varphi}_k\right)\right)^{-1}\mathbf{f}(\boldsymbol{\nu}_k) + \left(\mathbf{M}\left(\boldsymbol{\varphi}_{k+2}\right)\right)^{-1}\mathbf{f}(\boldsymbol{\nu}_{k+2}) - \left(\mathbf{M}\left(\boldsymbol{\varphi}_{k+1}\right)\right)^{-1}\mathbf{f}(\boldsymbol{\nu}_{k+1})$.

2) Piecewise changes of mean and constant covariance matrix: In this case $\mathbf{M}\left(\boldsymbol{\varphi}_j\right) = \mathbf{M}\left(\boldsymbol{\varphi}\right) = \boldsymbol{\Sigma}$, $\boldsymbol{\eta}_j = \left[\boldsymbol{\nu}_j^T,\ \boldsymbol{\varphi}^T\right]^T$, and $[\boldsymbol{\Phi}]_{kl}$ is given as follows:

For $\alpha_k > 0$, using Equation (33) and replacing $\mathbf{M}\left(\boldsymbol{\varphi}_k\right)$ and $\mathbf{M}\left(\boldsymbol{\varphi}_{k+1}\right)$ by $\boldsymbol{\Sigma}$, we have straightforwardly for $[\boldsymbol{\Phi}]_{kk}$:

$$[\boldsymbol{\Phi}]_{kk} = \exp\left\{\alpha_k\left(\mathbf{f}(\boldsymbol{\nu}_k) - \mathbf{f}(\boldsymbol{\nu}_{k+1})\right)^T\boldsymbol{\Sigma}^{-1}\left(\mathbf{f}(\boldsymbol{\nu}_k) - \mathbf{f}(\boldsymbol{\nu}_{k+1})\right)\right\}. \tag{36}$$

For $\alpha_k < 0$, using Equation (34), $[\boldsymbol{\Phi}]_{kk}$ is given as follows:

$$[\boldsymbol{\Phi}]_{kk} = \exp\left\{-\alpha_k\left(\mathbf{f}(\boldsymbol{\nu}_{k+1}) - \mathbf{f}(\boldsymbol{\nu}_k)\right)^T\boldsymbol{\Sigma}^{-1}\left(\mathbf{f}(\boldsymbol{\nu}_{k+1}) - \mathbf{f}(\boldsymbol{\nu}_k)\right)\right\}. \tag{37}$$

For $t_k + \alpha_k > t_{k+1} + \alpha_{k+1}$, using Equation (35), then$[\boldsymbol{\Phi}]_{kk+1}$is given as follows:

$$\begin{aligned}[\boldsymbol{\Phi}]_{kk+1} \quad = \quad & \left(\exp\left\{\frac{\beta_k}{2}\left(\left(\mathbf{f}(\boldsymbol{\nu}_{k+1}) - \mathbf{f}(\boldsymbol{\nu}_k)\right)\boldsymbol{\Sigma}^{-1}\left(\mathbf{f}(\boldsymbol{\nu}_{k+1}) - \mathbf{f}(\boldsymbol{\nu}_k)\right)^T\right.\right. \\ & \left.\left. + \left(\mathbf{f}(\boldsymbol{\nu}_{k+2}) - \mathbf{f}(\boldsymbol{\nu}_{k+1})\right)\boldsymbol{\Sigma}^{-1}\left(\mathbf{f}(\boldsymbol{\nu}_{k+2}) - \mathbf{f}(\boldsymbol{\nu}_{k+1})\right)^T - \left(\mathbf{f}(\boldsymbol{\nu}_k) - \mathbf{f}(\boldsymbol{\nu}_{k+2})\right)\boldsymbol{\Sigma}^{-1}\left(\mathbf{f}(\boldsymbol{\nu}_k) - \mathbf{f}(\boldsymbol{\nu}_{k+2})\right)^T\right)\right\}\right). \end{aligned} \tag{38}$$

3) Piecewise changes of covariance matrix and constant mean vector: In this case $\mathbf{f}(\boldsymbol{\nu}_j) = \mathbf{f}(\boldsymbol{\nu}) = \boldsymbol{\mu}$, $\boldsymbol{\eta}_j = \left[\boldsymbol{\nu}^T,\ \boldsymbol{\varphi}_j^T\right]^T$, and $[\boldsymbol{\Phi}]_{kl}$ is given as follows:

For $\alpha_k > 0$ using Equation (33) and replacing $\mathbf{f}(\boldsymbol{\nu}_k)$ and $\mathbf{f}(\boldsymbol{\nu}_{k+1})$ by $\boldsymbol{\mu}$, we have straightforwardly for $[\boldsymbol{\Phi}]_{kk}$:

$$[\boldsymbol{\Phi}]_{kk} = \begin{cases} \left(\frac{|\mathbf{M}(\boldsymbol{\varphi}_{k+1})|^{1/2}}{|\mathbf{M}(\boldsymbol{\varphi}_k)||\mathbf{M}_k|^{1/2}} \right)^{\alpha_k}, \text{ for } \mathbf{M}_k \in \mathbb{S}_{++}^M, \\ \infty, \text{ otherwise,} \end{cases} \tag{39}$$

where $\mathbf{M}_k = 2\left(\mathbf{M}\left(\boldsymbol{\varphi}_k\right)\right)^{-1} - \left(\mathbf{M}\left(\boldsymbol{\varphi}_{k+1}\right)\right)^{-1}$.

For $\alpha_k < 0$, using Equation (34), $[\boldsymbol{\Phi}]_{kk}$ is given as follows:

$$[\boldsymbol{\Phi}]_{kk} = \begin{cases} \left(\frac{|\mathbf{M}(\boldsymbol{\varphi}_k)|^{1/2}}{|\mathbf{M}(\boldsymbol{\varphi}_{k+1})||\mathbf{M}_{k+1}|^{1/2}} \right)^{-\alpha_k}, \text{ for } \mathbf{M}_{k+1} \in \mathbb{S}_{++}^M, \\ \infty, \text{ otherwise,} \end{cases} \tag{40}$$

where $\mathbf{M}_{k+1} = 2\left(\mathbf{M}\left(\boldsymbol{\varphi}_{k+1}\right)\right)^{-1} - \left(\mathbf{M}\left(\boldsymbol{\varphi}_k\right)\right)^{-1}$.

For $t_k + \alpha_k > t_{k+1} + \alpha_{k+1}$, using Equation (35), then$[\boldsymbol{\Phi}]_{kk+1}$is given as follows:

$$[\boldsymbol{\Phi}]_{kk+1} = \begin{cases} \left(\frac{|\mathbf{M}(\boldsymbol{\varphi}_{k+1})|^{1/2}}{|\mathbf{M}(\boldsymbol{\varphi}_k)|^{1/2}|\mathbf{M}(\boldsymbol{\varphi}_{k+2})|^{1/2}|\overline{\mathbf{M}}_k|^{1/2}} \right)^{\beta_k}, \quad \text{for } \overline{\mathbf{M}}_k \in \mathbb{S}_{++}^M, \\ \infty, \text{ otherwise,} \end{cases} \tag{41}$$

where $\overline{\mathbf{M}}_k = \left(\mathbf{M}\left(\boldsymbol{\varphi}_k\right)\right)^{-1} + \left(\mathbf{M}\left(\boldsymbol{\varphi}_{k+2}\right)\right)^{-1} - \left(\mathbf{M}\left(\boldsymbol{\varphi}_{k+1}\right)\right)^{-1}$.

The elements of Barankin bound for each case are obtained by using Equation (31), recalling that $A_k = [\boldsymbol{\Phi}]_{kk} - 1$ and $B_k = [\boldsymbol{\Phi}]_{kk+1} - 1$, from Equations (19) and (20), respectively.

B. Poisson case

Assume that the measurements $x_i \in \mathbb{N}+\{0\}$, for $i = 1, \ldots, N$, are distributed as a Poisson distribution with parameter $f(\boldsymbol{\eta}_j)$, where $f(\cdot)$ is a known function and $\boldsymbol{\eta}_j \in \mathbb{R}^L$ is a known parameter vector. Similarly to the Gaussian case, we restrict our analysis to the set of parameter vectors $\left\{\boldsymbol{\eta}_j\right\}$ such that the function $f(\boldsymbol{\eta}_j)$ is injective. Therefore, we derive closed-form expressions for the elements of matrix $\boldsymbol{\Phi} - \mathbf{1}_{q\times q}$ for piecewise changes of the parameter $\boldsymbol{\eta}_j$. Below, we evaluate $[\boldsymbol{\Phi}]_{kk}$ for $\alpha_k > 0$ and $\alpha_k < 0$, and $[\boldsymbol{\Phi}]_{kk+1}$ for $t_k + \alpha_k > t_{k+1} + \alpha_{k+1}$. Note that since $x_i \in \mathbb{N}$ we replace the integral operator by the summation operator.

For $\alpha_k > 0$, $[\boldsymbol{\Phi}]_{kk}$ becomes

$$[\boldsymbol{\Phi}]_{kk} = \exp\left\{ \frac{\alpha_k \left(f(\boldsymbol{\eta}_{k+1}) - f(\boldsymbol{\eta}_k)\right)^2}{f(\boldsymbol{\eta}_{k+1})} \right\}, \tag{42}$$

For $\alpha_k < 0$, $[\boldsymbol{\Phi}]_{kk}$ becomes

$$[\boldsymbol{\Phi}]_{kk} = \exp\left\{ \frac{-\alpha_k \left(f(\boldsymbol{\eta}_k) - f(\boldsymbol{\eta}_{k+1})\right)^2}{f(\boldsymbol{\eta}_k)} \right\}. \tag{43}$$

For $t_k + \alpha_k > t_{k+1} + \alpha_{k+1}$, $[\boldsymbol{\Phi}]_{kk+1}$ is given as follows:

$$[\boldsymbol{\Phi}]_{kk+1} = \exp\left\{ \beta_k \left(\frac{\left(f(\boldsymbol{\eta}_{k+1}) - f(\boldsymbol{\eta}_k)\right)^2}{2f(\boldsymbol{\eta}_{k+1})} + \frac{\left(f(\boldsymbol{\eta}_{k+2}) - f(\boldsymbol{\eta}_{k+1})\right)^2}{2f(\boldsymbol{\eta}_{k+1})} - \frac{\left(f(\boldsymbol{\eta}_k) - f(\boldsymbol{\eta}_{k+2})\right)^2}{2f(\boldsymbol{\eta}_{k+1})} \right) \right\}. \tag{44}$$

Similarly, as in the Gaussian case, the elements of Barankin bound for each case are obtained by using Equation (31) with $A_k = [\mathbf{\Phi}]_{kk} - 1$ and $B_k = [\mathbf{\Phi}]_{kk+1} - 1$.

V. Numerical Examples

In this section, as an illustration, we compare the MSE between the true values of the change-point locations and their maximum likelihood estimations with our bounds. In particular, we first introduce the MLE of change-point locations assuming the total number of changes is known. Then we analyze the cases of multiple changes in (i) the mean of a Gaussian distribution with fixed variance, (ii) the variance of a Gaussian assuming fixed mean, and (iii) the mean rate of Poisson distribution.

A. Maximum likelihood estimation

The MLE of $\mathbf{t}$ is the solution to the following problem:

$$\hat{\mathbf{t}}_{\text{ML}} = \arg\max_{\mathbf{t}} \sum_{i=1}^{q+1} \ln p_i(\mathbf{x}_{t_{i-1}+1}, \ldots, \mathbf{x}_{t_i}; \boldsymbol{\eta}_i), \tag{45}$$

where $t_0 = 0$ and $t_{q+1} = N$ by definition. There is no known closed-form expression for $\hat{\mathbf{t}}_{\text{ML}}$ so it has to be estimated via numerical computations. To solve this multidimensional optimization problem efficiently we apply dynamic programming (DP), explained in detail in [42], for our context of change-point estimation. The main advantage of the DP approach is that it does not need to evaluate all the possible combinations of values for $\mathbf{t}$ in (45). In all our examples below we illustrate the average MSE performance of the MLE for 1000 Monte Carlo experiments. We studied the performance as a function of signal-to-noise ratio (SNR), which is defined accordingly in each example, and as a function of the distance between change points. Here we chose $q = 3$ and the number of samples $N = 80$. In each example below, we set $t_2 = 40$, $t_3 = 60$, and we analyze two scenarios for change point t_1: In the first one, we set $t_1 = 20$ such that each segment has the same number of samples, and in the second scenario, $t_1 \in [2,\ 38]$.

Note that the unbiasedness properties of the MLE has been studied in [43] for a single change-point and for multiple change-points in [44]. The asymptotic results derived in [43] and [44] are applicable only for the case of a Gaussian distribution with changes in the mean. However, in the case of having a finite interval the MLE is expected to be biased independent of the distribution. On the other hand, it seems reasonable to assume that for large SNR values the MLE is approximately for a subset of the parameter space, i.e, subintervals, and specially for change-points located equidistant from their neighboring change-points or the interval limits. For example, in all the examples below, the bias of the MLE for $\mathbf{t} = [20\ 40\ 60]$ is approximately zero for all the SNR ranges considered in each scenario.

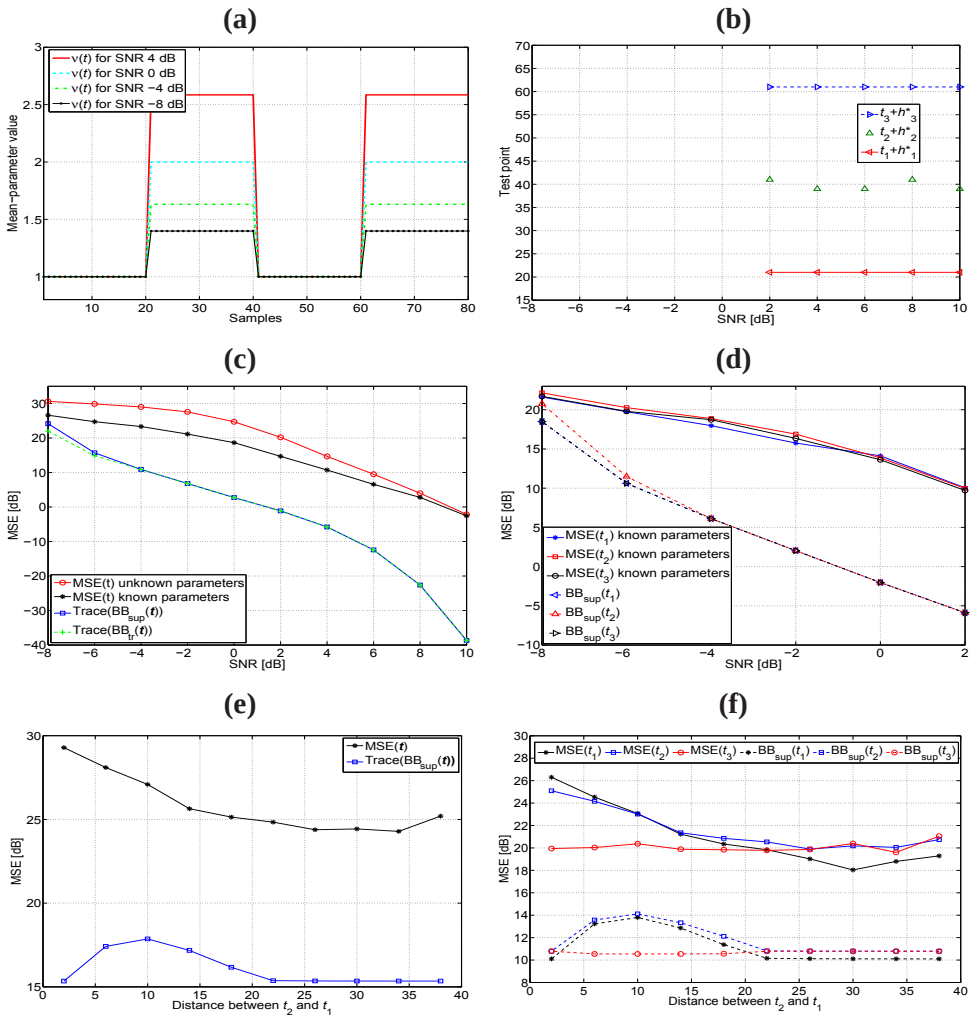

Fig. 1. Performance analysis for estimating change-points of the mean in a Gaussian distribution: (a) Mean values as a function of sample time for different SNR values; (b) Test points associated with the BB given by the minimal-upper bound of C, $\mathbf{BB}_{\sup}$, as a function of SNR; (c) MSE of the change-point vector using the MLE of $\mathbf{t}$ and its Barankin bound given by $\mathbf{BB}_{\sup}$, and by the matrix with maximum trace in C, $\mathbf{BB}_{\rm tr}$; (d) MSE of each change-point as a function of SNR using the MLE of t_1, t_2, and t_3 and their corresponding Barankin bound $\mathbf{BB}_{\sup}(t_i)$, $i = 1, \ldots, 3$; (e) MSE of change-point vector using the MLE of $\mathbf{t}$ and its Barankin bound, $\mathbf{BB}_{\sup}(\mathbf{t})$, as a function of the distance between t_2 and t_1 for SNR = -6 [dB]; (f) MSE of each change-point and their respective $\mathbf{BB}_{\sup}$ as a function of the distance between t_2 and t_1 for SNR = -6 [dB].

B. Changes in the mean of a Gaussian distribution

We consider the scenario of time series with 3 change points in the mean values of a Gaussian distribution with common variance. We recall the closed-form expressions obtained for computing $[\mathbf{\Phi}]_{kk}$, namely, Equations (36) and (37), and define the SNR for the k^{th} change point as

$$SNR_k = \left(\mathbf{f}(\boldsymbol{\nu}_{k+1}) - \mathbf{f}(\boldsymbol{\nu}_k)\right)^T \Sigma^{-1} \left(\mathbf{f}(\boldsymbol{\nu}_{k+1}) - \mathbf{f}(\boldsymbol{\nu}_k)\right), \tag{46}$$

where $\mathbf{f}(\boldsymbol{\nu}_k) \in \mathbb{R}^M$ is the mean vector of the k^{th} segment and $\boldsymbol{\Sigma} \in \mathbb{R}^{M\times M}$ is the common covariance matrix. In our example, $M = 1$ and, without loss of generality, we choose $f(\boldsymbol{\nu}_k) = \nu_k$ and $\Sigma = \sigma^2 = 1$, thus $\boldsymbol{\eta}_k = \left[\boldsymbol{\nu}_k^T,\ 1\right]^T$. Here, we set $\nu_1 = 1$ and ν_2, ν_3, and ν_4 are set such that $SNR_1 = SNR_2 = SNR_3 = SNR$. In particular, $\nu_k = \nu_{k-1} + (-1)^k \sqrt{\sigma^2 SNR}$ for $k = 2,\ 3,\ 4$. Figure 1(a) illustrates the mean values as a function of sample time for different SNR values. In Figure 1(c), we illustrate the MSE performance of the MLE for the change-point vector and the BB, as a function of SNR. In particular, $\text{MSE}_{\text{known}}$ is the MSE performance of the MLE for the change-point vector assuming knowledge of the means and variance, $\text{MSE}_{\text{unknown}}$ is the MSE performance of MLE for a more realistic case when no knowledge of the distribution parameter are available, $\mathbf{BB}_{\text{sup}}$ is given by the minimal-upper bound matrix $\mathbf{B}_q$ of the set C computed using the algorithm presented in Section III.D, and $\mathbf{BB}_{\text{tr}}$ is the matrix in C that has the maximum trace. We illustrate the trace of $\mathbf{BB}_{\text{sup}}$ and $\mathbf{BB}_{\text{tr}}$ since we are comparing the MSE performance for the change-point vector estimates. Note that, in view of the discussion presented in Section III.D, we compute $\mathbf{BB}_{\text{tr}}$ only in this example to show that $\mathbf{BB}_{\text{tr}}$ does not necessarily coincide with supremum of the set unless $\mathbf{BB}_{\text{sup}} \in C$. In this particular scenario, we found that $\mathbf{BB}_{\text{sup}}$ belongs to the set C for SNR values equal and larger than 2 dB. Therefore, we have optimal test points $\{\alpha_1^*,\ \alpha_2^*,\ \alpha_3^*\}$ associated to the matrix $\mathbf{BB}_{\text{sup}}$ defining the Lowner-John Ellipsoid , which are presented in Figure 1(b). For SNR values above 2 dB no change point is overlapped, therefore, each bound depends only on its corresponding diagonal element $[\boldsymbol{\Phi}]_{ii}$, which is equivalent to the resulting analysis of considering one change point located at $t = 20$ assuming $N = 40$. Moreover, it is important to mention that in this example, $[\boldsymbol{\Phi}]_{ii}$ is symmetric with respect to α_i and since all segments have the same length, then, both α_i and $-\alpha_i$ are optimal solutions for the bound on t_i. In Figure 1(b) we only illustrate one optimal solution. When the SNR > 2 dB we found the set C has several maximal elements that are not mutually comparable thus, $\mathbf{BB}_{\text{sup}} \notin C$ and does not show up in Figure 1(b). Finally, it can be seen that the test point approaches the true change point values as SNR increases; *i.e.*, α_2 tends to -1 as SNR increases.

In Figure 1(d), we illustrate the $\text{MSE}_{\text{known}}$ and $\mathbf{BB}_{\text{sup}}$ for change-point t_i, $i = 1, 2, 3$ as a function of SNR. It is noteworthy to mention that we did not illustrate the performance for higher SNR range in this example, since we found that for SNR values larger than 10 dB the bound tends quickly to zero. On the other hand, computing MSE values in these examples for larger SNR requires a large number of Monte Carlo simulations since the higher the SNR, the smaller the probability to have an error. For example, a single realization with an error of only 1 unit in one of the change-points, among 1000 realizations in the Monte Carlo simulation, amounts to an MSE of -30dB. Similar observations hold for the example on changes in the mean rate of a Poisson distribution.

We also analyze the MSE performance as a function of the distance between change points for a fixed SNR value. In Figure 1(e), for SNR $= -6$ dB, we illustrate the diagonal elements of $\mathbf{BB}_{\sup}$ and the MSE of the MLE for the change-point vector $\mathbf{t}$, assuming knowledge of the distribution parameters, as a function of the distance between change point t_1 and t_2. In Figure 1(f) we illustrate the BB and the MSE of the MLE for each change-point. We observe that the MSE of the MLE for t_1 and t_2 increases as the distance between change point t_1 and t_2 decreases. Similarly, their respective BB predict the same behavior for distances between t_1 and t_2 equal and larger than 10 time-units, however, for distances smaller than 10 time-units their respective bounds decrease to the same value as for distances larger than 22 time-units. This bound behavior is expected to take place as our Barankin-type lower bound approximation considers only one change-point per parameter. Therefore, in our problem the test-point values are lower and upper bounded by the adjacent change-point parameters, which does not allow for evaluating errors, in estimating each change-point, beyond these limits. Thus, as the change-points get closer the test-point domains become limited and the bound cannot take into account estimated errors given by estimates of t_1 which are larger than the true value of t_2, and estimated errors given by estimates of t_2 which are lower than the true value of t_1.

C. Changes in the variance of a Gaussian distribution

We consider the same scenarios as above, but with a time series with three change points in the variance of a Gaussian distribution and a common mean. We recall the closed-form expressions obtained for computing $[\mathbf{\Phi}]_{kk}$, namely, Equations (39) and (40), and define SNR for the k^{th} change point as $SNR_k = \frac{|\mathbf{M}(\boldsymbol{\varphi}_{k+1})|}{|\mathbf{M}(\boldsymbol{\varphi}_k)|}$, where $\mathbf{M}(\boldsymbol{\varphi}_k) \in \mathbb{R}^{M\times M}$ is the covariance matrix of the k^{th} segment. In our example, $M = 1$, and, without loss of generality, we choose $\mathbf{M}(\boldsymbol{\varphi}_k) = \varphi_k$ and the mean equal to zero since the BIM does not depend on the mean, thus $\boldsymbol{\eta}_k = [0,\ \varphi_k]^T$. Here, we set $\varphi_1 = 1$, and variances φ_2, φ_3, and φ_4 are set such that $SNR_1 = SNR_2 = SNR_3 = SNR$. In practice, $\varphi_k = \varphi_{k-1}SNR$. In Figure 2(a), we illustrate sigma-parameter values as a function of sample time for different SNR values. In Figure 2(c), we illustrate the MSE performance of the MLE for change-point vector as a function of SNR and its respective Barankin bound, $\mathbf{BB}_{\sup}$. In particular, we illustrate the $\text{MSE}_{\text{unknown}}$ and $\text{MSE}_{\text{known}}$ of $\mathbf{t}$ for SNR ranging from 1 to 30 dB. In Figure 2(d) we focus on the SNR ranging between 1 to 10 dB, and we illustrate the MSE for change-point estimate of t_1, t_2, and t_3 using MLE and their respective bounds given by the diagonal elements of $\mathbf{BB}_{\sup}$. In this scenario $\mathbf{BB}_{\sup}$ belongs to set C for SNR values larger than 4 dB, and the MSE of the MLE approaches slowly the BB as the SNR increases. In this example, the BB is the same for all change-points for SNR values above 2 dB,

and for all the SNR range illustrated, the maximum difference between the BB and both the $\text{MSE}_{\text{known}}$ and $\text{MSE}_{\text{unknown}}$ is approximately 7 dB and 17 dB, respectively. Note that for SNR values lower than 2 dB the $\mathbf{BB}_{\text{sup}}$ is greater than the MSE of the MLE, which is due to the fact that the Barankin bound derivation does not take into account the set of admissible values of the estimator. In our example, the MLE computation restricts the search to the range between 1 and N, and thus the MLE variance has an upper limit, which the BB computation does not consider. Moreover, note that BB assumes that the estimator is unbiased at the test-points, in addition to the parameter of interest, and since for low SNR values the optimal test-points tend to go to the extreme of the intervals associated to each change-point respectively, then the comparison against the BB tends to be inappropriate. Also, we illustrate in Figure 2(b), the optimal test points $[\alpha_1^*, \alpha_2^*, \alpha_3^*]^T$ associated to the matrix $\mathbf{BB}_{\text{sup}}$. It can be seen that for all the SNR range there are no overlaps between test points and, as in the previous example, all test points approach to 1 or -1, namely, they are close to the true change-point values as SNR increases. Therefore, for large SNR values $[\mathbf{BB}_{\text{sup}}]_{kk} = \frac{\sqrt{2\,SNR_k - 1}}{SNR_k}$, which tends to 0 as $SNR_k \to \infty$.

In Figures 2(e) and (f), for SNR $= 4$ dB, we illustrate the BB and the MSE of the MLE for t_1, t_2, and t_3, assuming knowledge of the distribution parameters, as a function of the distance between change point t_1 and t_2. The BB for all the change-points remains the same for distances between change-points t_1 and t_2 above 10 units. The BB for t_1 increases as the distance between change-points t_1 and t_2 increases from zero to 10 units. As in the previous example, the bound in this range is overly optimistic since the test-point domains become limited.

D. Changes in the mean rate of a Poisson distribution

Now we consider a time series with three change points in the mean rate of a Poisson distribution. Similarly as in the previous examples, we recall the closed-form expressions for $[\mathbf{\Phi}]_{kk}$, *i.e.*, Equations (42) and (43). Then we define SNR for the k^{th} change point detector as $SNR_k = \frac{\left(f(\boldsymbol{\eta}_k) - f(\boldsymbol{\eta}_{k+1})\right)^2}{f(\boldsymbol{\eta}_k)^2}$,where $f(\boldsymbol{\eta}_k)$ is the mean rate of the k^{th} segment. Here, without loss of generality, we set $f(\boldsymbol{\eta}_k) = \eta_k$. The mean rate is set $\eta_1 = 1$ and the mean rates η_2, η_3, and, η_4 are set such that $SNR_1 = SNR_2 = SNR_3 = SNR$. In practice, $\eta_k = \eta_{k-1}\left(1 + \sqrt{SNR}\right)$. In Figure 3(a), we illustrate the mean-rate-values as a function of sample time for different SNR values. Figures 3(c) and (d), illustrate the $\text{MSE}_{\text{unknown}}$ and $\text{MSE}_{\text{known}}$ performance for the change-point vector and each change points t_1, t_2, and t_3, respectively. In this case, the MSE values as well as the bounds for t_1, t_2, and t_3 are not the same for the same SNR values. In fact, it can be seen that the MSE values for t_3 are lower than the MSE values for t_2, and these last are lower than the MSE values for t_1. This difference in performance is due to the fact that in our example the

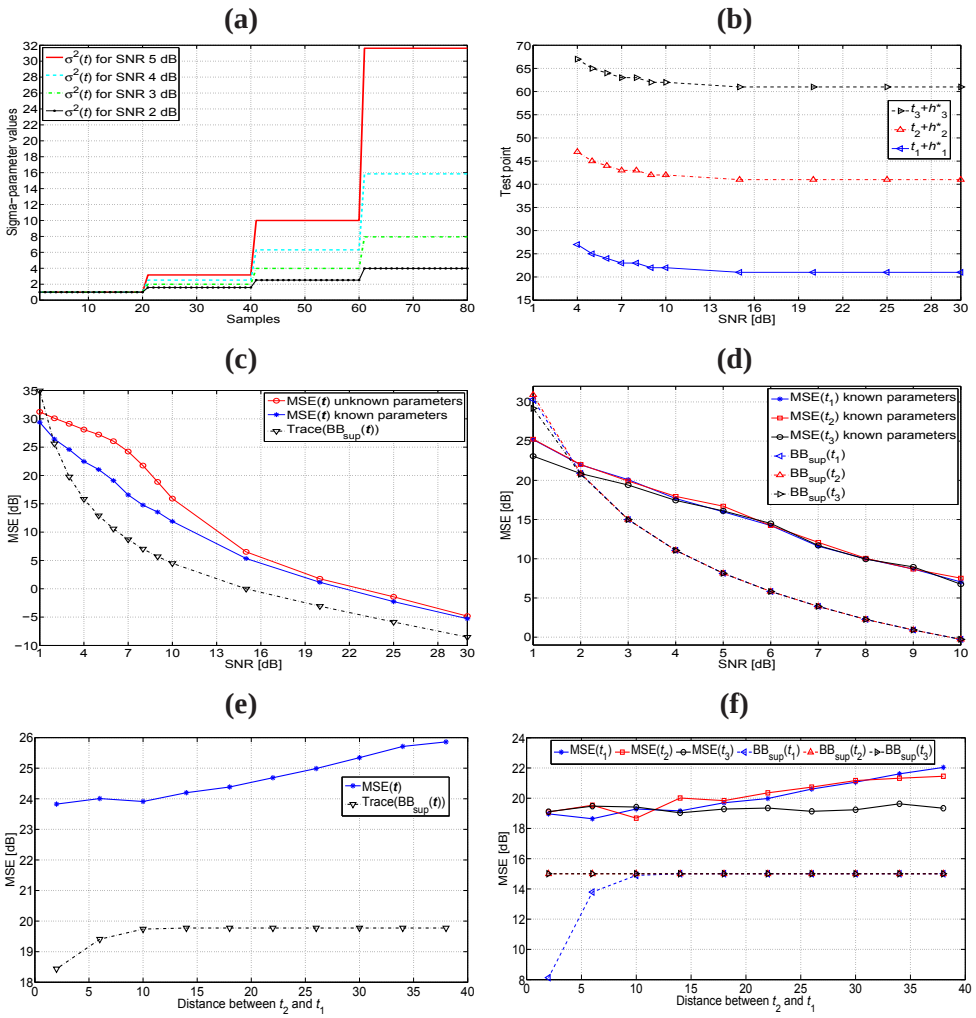

Fig. 2. Performance analysis for estimating change-points of the variance in a Gaussian distribution: (a) Sigma-parameter values as a function of sample time for different SNR values; (b) Test points associated with the BB given by the minimal-upper bound of C, $\mathbf{BB}_{\sup}$, as a function of SNR; (c) MSE of the change-point vector using the MLE of $\mathbf{t}$ and its Barankin bound given by $\mathbf{BB}_{\sup}$; (d) MSE of each change-point as a function of SNR using the MLE of t_1, t_2, and t_3 and their corresponding Barankin bound $\mathbf{BB}_{\sup}(t_i)$, $i = 1, \ldots, 3$; (e) MSE of change-point vector using the MLE of $\mathbf{t}$ and its Barankin bound, $\mathbf{BB}_{\sup}(\mathbf{t})$, as a function of the distance between t_2 and t_1 for SNR = $\mathbf{4}$ [dB]; (f) MSE of each change-point and their respective $\mathbf{BB}_{\sup}$ as a function of the distance between t_2 and t_1 for SNR = 4 [dB].

difference between the means of contiguous segments are not the same, which is a direct consequence of the definition used for SNR. In practice, for any SNR, the differences between the means for segments $[t_3 + 1,\ N]$ and $[t_2 + 1,\ t_3]$ is larger than the difference between the means for segments $[t_2 + 1,\ t_3]$ and $[t_1 + 1,\ t_2]$. In Figure 3(b) we illustrate the test points associated to the matrix $\mathbf{BB}_{\sup}$. As in the previous examples, the test points tend to the true change-point values as SNR increases. Finally, in Figures 3(e) and (f), we illustrate the MSE performance, assuming known mean rates, as a function of the distance

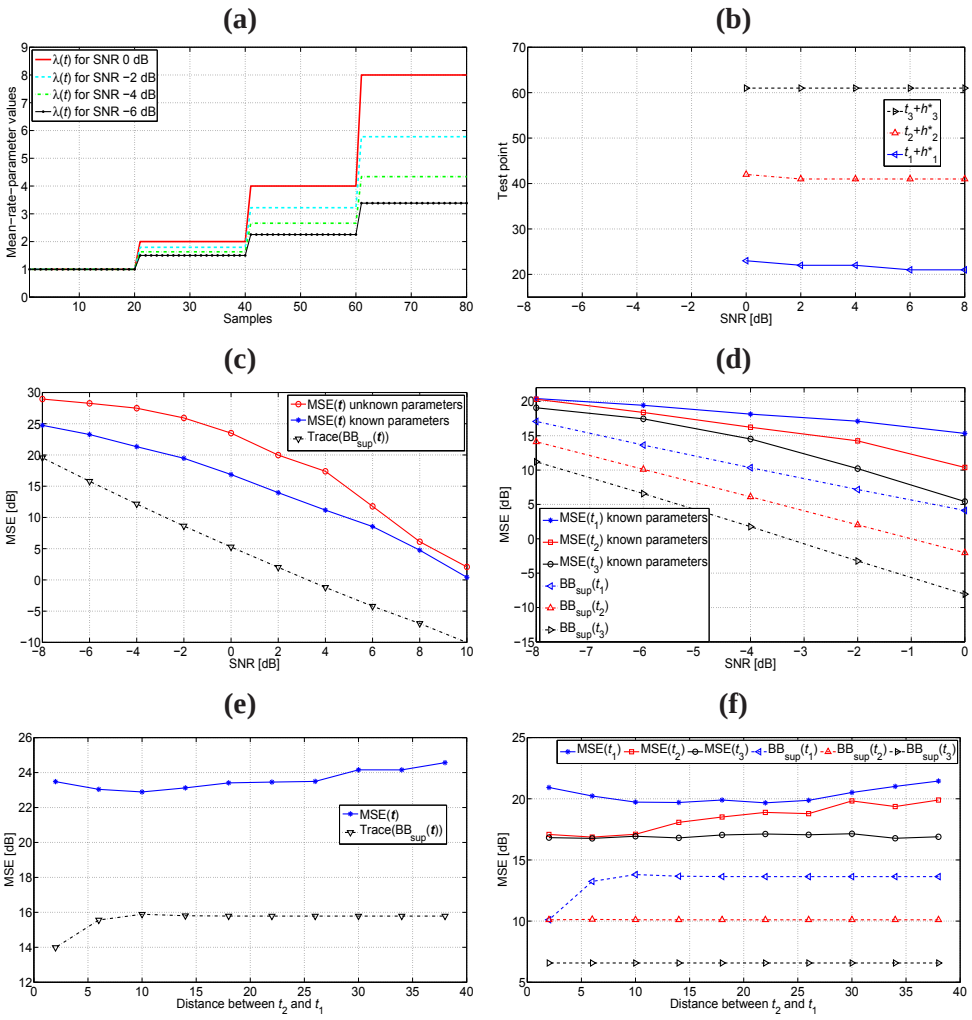

Fig. 3. Performance analysis for estimating change-points in the mean rate of a Poisson distribution distribution: (a) Mean-rate-values as a function of sample time for different SNR values; (b) Test points associated with the BB given by the minimal-upper bound of C, $\mathbf{BB}_{\sup}$, as a function of SNR; (c) MSE of the change-point vector using the MLE of $\mathbf{t}$ and its Barankin bound given by $\mathbf{BB}_{\sup}$; (d) MSE of each change-point as a function of SNR using the MLE of t_1, t_2, and t_3 and their corresponding Barankin bound $\mathbf{BB}_{\sup}(t_i)$, $i = 1, \ldots, 3$; (e) MSE of change-point vector using the MLE of $\mathbf{t}$ and its Barankin bound, $\mathbf{BB}_{\sup}(\mathbf{t})$, as a function of the distance between t_2 and t_1 for SNR = -6 [dB]; (f) MSE of each change-point and their respective $\mathbf{BB}_{\sup}$ as a function of the distance between t_2 and t_1 for SNR = -6 [dB].

between change points for SNR $= -6$ dB. The bounds for change-point t_2 and t_3 is constant in all the illustrated range, though, the MSE of the MLE for t_2 slightly varies as t_1 approaches t_2. As we discussed in the previous examples, the bound for t_1 is overly optimistic for small distance between t_2 and t_1 due to the constrained test-point domain.

VI. CONCLUSIONS

We investigated a simplified version of the Barankin bound on multiple change-point estimation. The approximate Barankin information matrix was spelled, revealing an interesting tri-diagonal structure, meaning that the estimation of one change point is naturally perturbed by its two neighbors. Moreover, the Barankin information matrix can be reduced to a block diagonal structure leading to closed-form for the elements of its inverse. The main limitation posed by this HCR approximation is a reduced search space for the BIM that leads to a loose Barankin bound. We also discussed the existence and computation of the supremum with respect to the Loewner partial ordering, on the finite set of candidate BB solutions. To overcome this problem, we computed a suitable minimal-upper bound to this set given by the matrix associated with the Lowner-John Ellipsoid of the set of hyper-ellipsoids associated to each maximal element of the set of candidate bound matrices. Two important distributions in signal and image processing were investigated, the Gaussian case and the Poisson case, for which we obtained closed-form expressions for all the elements of the Barankin information matrix. Finally, we illustrated our analysis by presenting various simulation results. In a future work, we will analyze Barankin-type lower bounds considering all distribution parameters in addition to the multiple change-point localizations.

APPENDIX

A. Proof of Lemma 1

Proof: We need to proof that for all $\mathbf{y} \in \mathbb{R}^q$ with $\mathbf{y} \neq \mathbf{0}$, $\mathbf{y}^T (\mathbf{A} - \mathbf{B}) \mathbf{y} > 0$ if $\lambda_1 \leq 1$. Since $\mathbf{A}$ is pd and $\mathbf{B}$ is psd, there exist a non-singular matrix $\mathbf{F}$ such that

$$\begin{aligned} \mathbf{F}^T \mathbf{B} \mathbf{F} &= \mathbf{diag}\left(\lambda_1, \ldots, \lambda_m, \lambda_{m+1}, \ldots, \lambda_q\right) = \Lambda \text{ and} \\ \mathbf{F}^T \mathbf{A} \mathbf{F} &= \mathbf{I} \end{aligned}$$

Thus, $\mathbf{B} = \left(\mathbf{F}^T\right)^{-1} \Lambda \left(\mathbf{F}\right)^{-1}$ and $\mathbf{A} = \left(\mathbf{F}^T\right)^{-1} \mathbf{I} \left(\mathbf{F}\right)^{-1}$and $\mathbf{y}^T (\mathbf{A} - \mathbf{B}) \mathbf{y} = \mathbf{y}^T \left(\mathbf{F}^T\right)^{-1} (\mathbf{I} - \Lambda) \left(\mathbf{F}\right)^{-1} \mathbf{y}$. Let $\mathbf{z} = \left(\mathbf{F}\right)^{-1} \mathbf{y}$, because $\mathbf{F}$ is not singular $\left(\mathbf{F}\right)^{-1} \mathbf{y} = \mathbf{0}$ for $\mathbf{y} = \mathbf{0}$, therefore our problem is equivalent to analyze the positiveness of $r = \mathbf{z}^T (\mathbf{I} - \Lambda) \mathbf{z}$, for $\mathbf{z} \neq \mathbf{0}$. Since $\lambda_{m+1} = \cdots = \lambda_q = 0$, $r = \sum_{i=1}^{m} (1 - \lambda_i) z_i^2 + \sum_{i=m+1}^{q} z_i^2$. Hence, if $\lambda_1 \leq 1$, then $(1 - \lambda_i) \geq 0$, for $i = 1, \ldots, m$ thus $r > 0$ and $\mathbf{A} \succ \mathbf{B}$. On the other hand, if $\lambda_1 > 1$, we can always find a $\mathbf{z}$ vector such that $r \leq 0$ or $r > 0$, thus $\mathbf{A}$ and $\mathbf{B}$ are not mutually comparable. ■

B. Computing diagonal elements of $\mathbf{\Phi}$

For $\alpha_k > 0$, Equation (12) becomes

$$[\Phi]_{kk} = \int_{\Omega} \prod_{i=1}^{t_1} p_1(x_i;\eta_1)\cdots \prod_{i=t_{k-1}+1}^{t_k} p_k(x_i;\eta_k) \frac{\prod_{i=t_k+1}^{t_k+\alpha_k} p_k^2(x_i;\eta_k)}{\prod_{i=t_k+1}^{t_k+\alpha_k} p_{k+1}(x_i;\eta_{k+1})} \prod_{i=t_k+\alpha_k+1}^{t_{k+1}} p_{k+1}(x_i;\eta_{k+1})\cdots \prod_{i=t_q+1}^{N} p_{q+1}(x_i;\eta_{q+1}) dX.$$

After some straightforward simplifications, we have that

$$[\Phi]_{kk} = \int_{\Omega} \prod_{i=t_k+1}^{t_k+\alpha_k} \frac{p_k^2(\mathbf{x}_i;\boldsymbol{\eta}_k)}{p_{k+1}(\mathbf{x}_i;\boldsymbol{\eta}_{k+1})} dx_{t_k+1}\cdots dx_{t_k+\alpha_k} = \left(\int_{\Omega} \frac{p_k^2(\mathbf{x};\boldsymbol{\eta}_k)}{p_{k+1}(\mathbf{x};\boldsymbol{\eta}_{k+1})} d\mathbf{x}\right)^{\alpha_k}.$$

Similar analysis can be applied to solve for $\alpha_k < 0$.

C. Computing non-diagonal elements of $\mathbf{\Phi}$

For $\alpha_k > 0$ and $\alpha_l > 0$, Equation (14) becomes

$$\begin{aligned}[\mathbf{\Phi}]_{kl} = \int_{\Omega} & \frac{\prod_{i=t_k+1}^{t_k+\alpha_k} p_k(\mathbf{x}_i;\boldsymbol{\eta}_k)}{\prod_{i=t_k+1}^{t_k+\alpha_k} p_{k+1}(\mathbf{x}_i;\boldsymbol{\eta}_{k+1})} \prod_{i=1}^{t_1} p_1(\mathbf{x}_i;\boldsymbol{\eta}_1)\cdots \prod_{i=t_k+1}^{t_k+\alpha_k} p_{k+1}(\mathbf{x}_i;\boldsymbol{\eta}_{k+1}) \prod_{i=t_k+\alpha_k+1}^{t_{k+1}} p_{k+1}(\mathbf{x}_i;\boldsymbol{\eta}_{k+1}) \\ & \cdots \prod_{i=t_{l-1}+1}^{t_l+\alpha_l} p_l(\mathbf{x}_i;\boldsymbol{\eta}_l)\cdots \prod_{i=t_q+1}^{N} p_{q+1}(\mathbf{x}_i;\boldsymbol{\eta}_{q+1}) d\mathbf{X} = 1.\end{aligned}$$

The cases $(\alpha_k < 0,\ \alpha_l < 0)$, $(\alpha_k < 0,\ \alpha_l > 0)$, and $t_k + \alpha_k < t_l + \alpha_l$ are solved using same approach as above. For the overlapping case, i.e., $t_k + \alpha_k > t_l + \alpha_l$, is more difficult. Replacing $l = k+1$ and keeping in mind that $\alpha_k > 0$ and $\alpha_{k+1} < 0$, Equation (14) becomes

$$\begin{aligned}[\mathbf{\Phi}]_{kk+1} &= \int_{\Omega} \prod_{i=1}^{t_1} p_1(\mathbf{x}_i;\boldsymbol{\eta}_1) \prod_{i=t_{k+1}+\alpha_{k+1}+1}^{t_k+\alpha_k} \frac{p_k(\mathbf{x}_i;\boldsymbol{\eta}_k) p_{k+2}(\mathbf{x}_i;\boldsymbol{\eta}_{k+2})}{p_{k+1}(\mathbf{x}_i;\boldsymbol{\eta}_{k+1})} \cdots \prod_{i=t_q+1}^{N} p_{q+1}(\mathbf{x}_i;\boldsymbol{\eta}_{q+1}) d\mathbf{X} \\ &= \left(\int_{\Omega} \frac{p_k(\mathbf{x};\boldsymbol{\eta}_k) p_{k+2}(\mathbf{x};\boldsymbol{\eta}_{k+2})}{p_{k+1}(\mathbf{x};\boldsymbol{\eta}_{k+1})} d\mathbf{x}\right)^{\beta_k},\end{aligned}$$

where $\beta_k = (t_k + \alpha_k) - (t_{k+1} + \alpha_{k+1})$.

D. Computing the elements of $\mathbf{\Phi}$ *for changes in mean and covariance matrix of Gaussian distribution*

In this case $\boldsymbol{\eta}_j = \left[\boldsymbol{\nu}_j^T,\ \boldsymbol{\varphi}_j^T\right]^T$, and the data likelihood is given as follows,

$$p(\mathbf{X};\mathbf{t}) = \frac{1}{(2\pi)^{N\ M/2} \prod_{j=1}^{q+1} \left|\mathbf{M}(\boldsymbol{\varphi}_j)\right|^{(t_j - t_{j-1})/2}} \times \exp\left\{-\frac{1}{2} Tr\left\{\sum_{j=1}^{q+1} \mathbf{M}(\boldsymbol{\varphi}_j)^{-1} \left(\sum_{i=t_{j-1}+1}^{t_j} \left(\mathbf{x}_i - \mathbf{f}(\boldsymbol{\nu}_j)\right)\left(\mathbf{x}_i - \mathbf{f}(\boldsymbol{\nu}_j)\right)^T\right)\right\}\right\}.$$

For $\alpha_k > 0$, using Equation (13), we have that $[\mathbf{\Phi}]_{kk}$ is given as follows:

$$\begin{aligned}[\mathbf{\Phi}]_{kk} = & \left(\frac{\left|\mathbf{M}(\boldsymbol{\varphi}_{k+1})\right|^{1/2}}{(2\pi)^{M/2}\left|\mathbf{M}(\boldsymbol{\varphi}_k)\right|}\right)^{\alpha_k} \left(\int_{\mathbb{R}^M} \exp\left\{-\frac{1}{2}\left(\mathbf{x}_i^T \mathbf{M}_k \mathbf{x}_i - 2\mathbf{g}_k^T \mathbf{x}_i\right)\right\} d\mathbf{x}_i\right)^{\alpha_k} \\ & \times \exp\left\{-\frac{\alpha_k}{2}\mathbf{f}^T(\boldsymbol{\nu}_k) 2\left(\mathbf{M}(\boldsymbol{\varphi}_k)\right)^{-1}\mathbf{f}(\boldsymbol{\nu}_k) + \frac{\alpha_k}{2}\mathbf{f}^T(\boldsymbol{\nu}_{k+1})\left(\mathbf{M}(\boldsymbol{\varphi}_{k+1})\right)^{-1}\mathbf{f}(\boldsymbol{\nu}_{k+1})\right\},\end{aligned}$$

where $M_k = \left(2\left(\mathbf{M}\left(\boldsymbol{\varphi}_k\right)\right)^{-1} - \left(\mathbf{M}\left(\boldsymbol{\varphi}_{k+1}\right)\right)^{-1}\right)$ and $g_k = 2\left(\mathbf{M}\left(\boldsymbol{\varphi}_k\right)\right)^{-1} f(\nu_k) - \left(\mathbf{M}\left(\boldsymbol{\varphi}_{k+1}\right)\right)^{-1} f(\nu_{k+1})$. The integral above has a finite value for $\mathbf{M}_k$ positive definite (pd). Hence, and after some straightforward algebraic derivations, we obtain the expression in (33). The case $\alpha_k < 0$ is obtained proceeding similarly as above. For $t_k + \alpha_k > t_{k+1} + \alpha_{k+1}$, using Equation (17), we have that $[\boldsymbol{\Phi}]_{kk+1}$is given as follows:

$$
\begin{aligned}
[\boldsymbol{\Phi}]_{kk+1} \quad = \quad & \left(\frac{\left|\mathbf{M}\left(\boldsymbol{\varphi}_{k+1}\right)\right|^{1/2}}{(2\pi)^{M/2}\left|\mathbf{M}\left(\boldsymbol{\varphi}_k\right)\right|^{1/2}\left|\mathbf{M}\left(\boldsymbol{\varphi}_{k+2}\right)\right|^{1/2}}\right)^{\beta_k}\left(\int_{\mathbb{R}^M}\exp\left\{-\frac{1}{2}\left(\mathbf{x}_i^T\overline{\mathbf{M}}_k\mathbf{x}_i - 2\overline{\mathbf{g}}_k^T\mathbf{x}_i\right)\right\}d\mathbf{x}\right)^{\beta_k} \\
& \times\exp\left\{-\frac{\beta_k}{2}\mathbf{f}^T(\boldsymbol{\nu}_k)\left(\mathbf{M}\left(\boldsymbol{\varphi}_k\right)\right)^{-1}\mathbf{f}(\boldsymbol{\nu}_k) - \frac{\beta_k}{2}\mathbf{f}^T(\boldsymbol{\nu}_{k+2})\left(\mathbf{M}\left(\boldsymbol{\varphi}_{k+2}\right)\right)^{-1}\mathbf{f}(\boldsymbol{\nu}_{k+2}) + \frac{\beta_k}{2}\mathbf{f}^T(\boldsymbol{\nu}_{k+1})\left(\mathbf{M}\left(\boldsymbol{\varphi}_{k+1}\right)\right)^{-1}\mathbf{f}(\boldsymbol{\nu}_{k+1})\right\}
\end{aligned}
$$

where $\overline{\mathbf{M}}_k = \left(\mathbf{M}\left(\boldsymbol{\varphi}_k\right)\right)^{-1} + \left(\mathbf{M}\left(\boldsymbol{\varphi}_{k+2}\right)\right)^{-1} - \left(\mathbf{M}\left(\boldsymbol{\varphi}_{k+1}\right)\right)^{-1}$, and $\overline{\mathbf{g}}_k = \left(\mathbf{M}\left(\boldsymbol{\varphi}_k\right)\right)^{-1}\mathbf{f}(\boldsymbol{\nu}_k) + \left(\mathbf{M}\left(\boldsymbol{\varphi}_{k+2}\right)\right)^{-1}\mathbf{f}(\boldsymbol{\nu}_{k+2}) - \left(\mathbf{M}\left(\boldsymbol{\varphi}_{k+1}\right)\right)^{-1}\mathbf{f}(\boldsymbol{\nu}_{k+1})$. Hence, and after some straightforward algebraic derivations, we obtain the expression in (35).

REFERENCES

[1] A. Benveniste and M. Basseville, *Detection of Abrupt Changes in Signals and Dynamical Systems*. Springer-Verlag, Berlin, 1986.

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

[3] F. Gustafsson, *Adaptive Filtering and Change Detection*, 1st ed. Wiley, Oct. 2000.

[4] B. Brodskya and B. Darkhovskyb, *Non-Parametric Statistical Diagnosis. Problems and Methods*. Kluwer Academic Publishers Dordrecht/Boston/London, 2000.

[5] ——, "Asymptotically optimal methods of change-point detection for composite hypotheses," *Journal of Statistical Planning and Inference*, vol. 133, pp. 123–138, 2005.

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

[7] R. Kakarala and A. O. Hero, "On achievable accuracy in edge localization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 7, pp. 777–781, July 1992.

[8] A. Bartov and H. Messer, "Analysis of inherent limitations in localizing step-like singularities in a continuous signal," in *Proc. of the IEEE-SP Intl. Symposium on Time-Frequency and Time-Scale Analysis*, Paris, FR, June 1996, pp. 21–24.

[9] A. M. Reza and M. Doroodchi, "Cramér-Rao lower bound on locations of sudden changes in a steplike signal," *IEEE Transactions on Signal Processing*, vol. 44, no. 10, pp. 2551–2556, Oct. 1996.

[10] J.-Y. Tourneret, M. Chabert, and M. Ghogho, "Detection and estimation of multiplicative jumps," in *Proc. of the 8th IEEE Signal Processing Workshop on Statistical Signal and Array Processing*, June 1996, pp. 20–23.

[11] A. Bartov and H. Messer, "Lower bound on the achievable DSP performance for localizing step-like continuous signals in noise," *IEEE Transactions on Signal Processing*, vol. 46, no. 8, pp. 2195–2201, Aug. 1998.

[12] A. Swami and B. Sadler, "Analysis of multiscale products for step detection and estimation," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 1043–1051, Apr. 1999.

[13] S. Zacks, *The Theory of Statistical Inference*. J. Wiley & Sons, N. York, 1971.

[14] A. Swami and B. Sadler, "Cramér-Rao bounds for step-change localization in additive and multiplicative noise," in *Proc. of the IEEE Signal Processing Workshop on Statistical Signal and Array Processing*, Sept. 1999, pp. 403–406.

[15] J. V. Braun and H. G. Muller, "Statistical methods for DNA sequence segmentation," *Statistical Science*, vol. 13, no. 2, pp. 142–162, 1998.

[16] P. S. La Rosa, A. Nehorai, H. Eswaran, C. Lowery, and H. Preissl, "Detection of uterine MMG contractions using a multiple change point estimator and the K-means cluster algorithm," *Transactions of Biomedical Engineering*, vol. 55, no. 2, pp. 453–467, Feb. 2008.

[17] C. H. Wu and C. H. Hsieh, "Multiple change-point audio segmentation and classification using an MDL-based Gaussian model," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 2, pp. 647–657, Mar. 2006.

[18] N. Dobigeon, J.-Y. Tourneret, and J. D. Scargle, "Change-point detection in astronomical data by using a hierarchical model and a Bayesian sampling approach," in *IEEE/SP 13th Workshop on Statistical Signal Processing*, July 2005, pp. 369–374.

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

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

[21] L. Knockaert, "The Barankin bound and threshold behaviour in frequency estimation," *IEEE Transactions on Signal Processing*, vol. 45, no. 9, pp. 2398–2401, Sept. 1997.

[22] A. Ferrari and J.-Y. Tourneret, "Barankin lower bound for change-points in independent sequences," in *Proc. of IEEE Workshop on Statistical Signal Processing (SSP)*, St. Louis, MO, USA, Sept. 2003, pp. 557–560.

[23] J. M. Hammersley, "On estimating restricted parameters," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 12, no. 2, pp. 192–240, 1950.

[24] B. Burgeth, A. Bruhna, N. Papenberg, M. Welka, and J. Weickert, "Mathematical morphology for matrix fields induced by the Loewner ordering in higher dimensions," *Signal Process.*, vol. 87, no. 2, pp. 277–290, Feb. 2007.

[25] F. Pukelsheim, *Optimal Design of Experiments*, 1st ed. John Wiley & Sons Inc., 1993.

[26] J. Borwein and A. Lewis, *Convex Analysis and Nonlinear Optmization: Theory and Examples*, 1st ed. Springer-Verlag, 2000.

[27] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Transactions on Information Theory*, vol. 26, no. 6, Nov. 1990.

[28] 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.

[29] R. J. McAulay and L. P. Seidman, "A useful form of the Barankin lower bound and its application to PPM threshold analysis," *IEEE Transactions on Information Theory*, vol. 15, pp. 273–279, Mar. 1969.

[30] R. J. McAulay and E. M. Hofstetter, "Barankin bounds on parameter estimation," *IEEE Transactions on Information Theory*, vol. 17, pp. 669–676, Nov. 1971.

[31] J. S. Abel, "A bound on mean square estimate error," *IEEE Transactions on Information Theory*, vol. 39, no. 5, pp. 1675–1680, Sept. 1993.

[32] J. Tabrikian and J. Krolik, "Barankin bounds for source localization in an uncertain ocean environment," *IEEE Transactions on Signal Processing*, vol. 47, no. 11, pp. 2917–2927, Nov. 1999.

[33] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[34] M. Grant and S. Boyd, "Cvx: Matlab software for disciplined convex programming (web page and software)," *http://stanford.edu/ boyd/cvx*, December 2008.

[35] ——, "Graph implementations for nonsmooth convex programs," *Recent Advances in Learning and Control (a tribute to M. Vidyasagar)*, December 2008.

[36] G. Strang, *Introduction to Linear Algebra*. Wellesley, MA : Wellesley-Cambridge Press, 2005.

[37] T. Muir, *A treatise on the theory of determinants*. Dover Publications, 1960.

[38] P. Swerling, "Probability of detection for fluctuating targets," *IEEE Transactions on Information Theory*, vol. 6, no. 2, pp. 269–308, Apr. 1960.

[39] J. F. Böhme, "Estimation of source parameters by maximum likelihood and non linear regression," in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 1984, pp. 731–734.

[40] ——, "Estimation of spectral parameters of correlated signals in wavefields," *Signal Process.*, vol. 10, pp. 329–337, 1986.

[41] Z. Liu, A. Nehorai, and E. Paldi, "Statistical analysis of a generalized compound eye detector array," in *Proc. 4th IEEE Sensor Conference*, Oct.-Nov. 2005.

[42] S. M. Kay, *Fundamentals of Statistical Signal Processing, Vol. 2: Detection Theory*, 1st ed. Prentice Hall PTR, Jan. 1998.

[43] D. V. Hinkley, "Inference about the change-point in a sequence of random variables," *Biometrika*, vol. 57, no. 1, pp. 1–17, 1970.

[44] Y.-C. Yao and S. T. Au, "Least-squares estimation of a step function," *Sankhya: The Indian Journal of Statistics, Series A*, vol. 51, no. 3, pp. 370–381, October 1989.