Prior Influence on Weiss-Weinstein Bounds for Multiple Change-Point Estimation

Lucien Bacharach*, Alexandre Renaux* and Mohammed Nabil El Korso†
*Université Paris-Sud/L2S
†Université Paris-Nanterre/LEME

***Abstract*—In this paper, we study the influence of the prior distribution on the Weiss-Weinstein bound (WWB) in the ubiquitous multiple change-point estimation problem. As by product, we derive a closed form expressions of the WWB for different commonly used prior. Numerical results reveal some insightful properties and corroborate the proposed theoretical analysis.**

I. Introduction

Lower bounds for the mean square error (MSE) are commonly used to assess the estimation performance for a given problem. The most widely used lower bound is the Cramér-Rao bound (CRB) since it is known to be a tight bound under some regularity conditions [1]. One of these regularity conditions is the differentiability of the log-likelihood function. In the estimation of discrete parameters context, the CRB does not exist, which requires the use of other lower bounds [2].

An important example of discrete parameter estimation problem is the change-point problem, in which the distribution of the observations abruptly changes at an unknown time instant, named "change-point". Other lower bounds, then the CRB, with relaxed regularity conditions have been derived for the change-point problem, as Barankin-type lower bounds [3], in the case of a single change [4] and later extended to multiple changes [5], yielding coarse approximations of the change-point estimation behavior. The use of the Bayesian point of view, in which the unknown change-point location is assumed to be a random parameter, allows to derive the Weiss-Weinstein bound (WWB) which is generally tight, even for low information regimes [2], [6]. In such low information conditions, as for instance under low signal-to-noise ratio (SNR), prior plays an important role, and for this reason, it is interesting to assess its influence on the WWBs. The prior influence has been studied for linear models in [7]. In this paper, we propose to study this prior influence in the context of multiple change-points. A WWB for this problem was recently derived using a particular uniform random walk as the prior [8]. We extend this work here to a wider class of prior distribution for the change locations.

II. Problem set-up

In this section, we first present the signal model and the associated estimation problem we consider in this paper. Using the Bayesian point of view, we then present the different prior distributions investigated for this problem.

A. The multiple change-point problem

We consider a classical multiple change-point framework, where a time series $\boldsymbol{x} \triangleq [x_1, \ldots, x_N]$ made up of N independent observations is separated into $Q+1$ independent segments, delineated by Q unknown change-points $t_1, \ldots, t_Q$:

$$\begin{cases} x_n \sim p_{\boldsymbol{\eta}_1}(x_n) & \text{for } n = 1, \ldots, t_1 \\ \quad\vdots & \quad\quad\vdots \\ x_n \sim p_{\boldsymbol{\eta}_{Q+1}}(x_n) & \text{for } n = t_Q + 1, \ldots, N \end{cases} \tag{1}$$

For this model, defining the unknown parameter vector $\boldsymbol{t} \triangleq [t_1, \ldots, t_Q]$, the likelihood of the observations can be written as $p(\boldsymbol{x} \,|\, \boldsymbol{t}) = \prod_{q=1}^{Q+1} \prod_{n=t_q+1}^{t_{q+1}} p_{\boldsymbol{\eta}_q}(x_n)$.

B. Various priors for the Bayesian approach

From the Bayesian setting, the vector $\boldsymbol{t}$ is assigned a prior distribution $\pi(\boldsymbol{t})$. In the change-point context, when changes are known to occur in the observation window, the support of the prior distribution, denoted by $\mathscr{T}$, is finite. Here, we assume that there are Q changes, thus we have $\mathscr{T} \subseteq \{\boldsymbol{t} \in \mathbb{N}^Q \,|\, 0 < t_1 < \ldots < t_Q < N\}$. The joint distribution of the observations and the parameter vector $\boldsymbol{t}$ is denoted by $p(\boldsymbol{x}, \boldsymbol{t}) \triangleq p(\boldsymbol{x} \,|\, \boldsymbol{t})\, \pi(\boldsymbol{t})$. Let us now enumerate the various prior distributions considered throughout this paper, and give their expressions.

1) Uniform on independent segments (UIS) distribution: This type of prior distribution is defined as

$$\pi^{\mathsf{uis}}(\boldsymbol{t}) = \prod_{q=1}^{Q} \frac{\mathbb{1}_{[\![\tau_{q-1}+1, \tau_q]\!]}(t_q)}{\tau_q - \tau_{q-1}} \tag{2}$$

where $\tau_0 \triangleq 0$, and the $(\tau_q)_{q=1,\ldots,Q}$ can be freely chosen, as long as $0 < \tau_1 < \ldots < \tau_Q < N$. For instance, we can choose $\tau_q = q\lfloor (N-1)/Q \rfloor$ (where $\lfloor . \rfloor$ denotes the floor function) to promote segments with uniform lengths, and distributed all along the observation window.

2) Uniform random walk (URW): $t_q = t_{q-1} + \epsilon_q$ ($q = 1, \ldots, Q$, $t_0 \triangleq 0$), where ϵ_q is distributed according to a uniform distribution $\mathcal{U}([\![d\,, D]\!])$, with d and D some freely chosen integers satisfying $1 \leq d < D \leq \lfloor N/(Q+1) \rfloor$ (since we assumed the total number of changes equals Q).

$$\pi^{\mathsf{urw}}(\boldsymbol{t}) = \prod_{q=1}^{Q} \frac{1}{\Delta} \mathbb{1}_{[\![t_{q-1}+d, t_{q-1}+D]\!]}(t_q) \tag{3}$$

where we have defined $\Delta \triangleq D - d + 1$. This prior was used in [8].

ISBN 978-90-827970-1-5 © EURASIP 2018

3) Globally uniform (GU) distribution:

$$\pi^{\mathsf{gu}}(\boldsymbol{t}) = \prod_{q=1}^{Q} \frac{1}{\binom{N-1}{Q}} \mathbb{1}_{[\![t_{q-1}+1, N-Q+q-1]\!]}(t_q). \quad (4)$$

Notice that is the most uninformative distribution among those considered in this paper. This prior was used in [9] for instance.

4) Truncated Poisson random walk (PRW): Finally, the last prior distribution investigated in this paper is a truncated Poisson random walk. This means that each change-point t_q, given the previous one t_{q-1}, can be written as $t_q = t_{q-1}+1+u_q$, $q = 1, \ldots, Q$, where u_q follows a truncated Poisson distribution with parameter λ. This parameter λ actually denotes the *a priori* mean length (minus 1) of the segments $[t_{q-1}\,, t_q]$ in the time series. Consequently, we set $\lambda = N/(Q+1) - 1$. Hence, the joint *a priori* distribution of the change-points is given by

$$\begin{aligned}\pi^{\mathsf{prw}}(\boldsymbol{t}) &= \prod_{q=1}^{Q} \Pr(T_q = t_q \,|\, T_{q-1} = t_{q-1} \text{ and} \\ &\qquad\qquad T_{q-1}+1 \le T_q \le N-Q+q-1) \\ &= \frac{\lambda^{t_Q - Q}\, \mathbb{1}_{[\![t_{q-1}+1, N-Q+q-1]\!]}(t_q)}{\prod_{q=1}^{Q} \left((t_q - t_{q-1} - 1)! \sum_{i=0}^{N-Q+q-2-t_{q-1}} \frac{\lambda^i}{i!} \right)} \end{aligned} \quad (5)$$

in which the capitalized versions T_{q-1}, T_q, U_q of t_{q-1}, t_q, u_q denote the associated random variables (note: these random variables are omitted elsewhere in the paper for sake of shortness). This prior distribution is inspired from that used in [10], and adapted to the case of a known number Q of changes (hence, the truncation).

III. Derivation of the WWB for the multiple change-point problem with various priors

A. Background on the WWB

For any Bayesian estimator $\hat{\boldsymbol{t}}(\boldsymbol{x})$ of a parameter vector $\boldsymbol{t} \triangleq [t_1, \ldots, t_Q] \in \mathcal{T} \subset \mathbb{R}^Q$, based on the observations $\boldsymbol{x}$, the Weiss-Weinstein bound is a lower bound for the mean square error of $\hat{\boldsymbol{t}}(\boldsymbol{x})$. For $Q \ge 2$, the WWB is a matrix $\boldsymbol{W}(\boldsymbol{H})$ parameterized by the so-called "test-point matrix" $\boldsymbol{H} \triangleq [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_Q]$, whose column vectors $\boldsymbol{h}_q$, $q = 1, \ldots, Q$ (the "test-points") belong to $\mathcal{H} \triangleq \{\boldsymbol{h} \in \mathbb{R}^Q \,|\, p(\boldsymbol{x}, \boldsymbol{t}+\boldsymbol{h}) > 0\}$. According to [2], the matrix difference $\mathbb{E}_{\boldsymbol{x},\boldsymbol{t}}\{[\hat{\boldsymbol{t}}(\boldsymbol{x}) - \boldsymbol{t}][\hat{\boldsymbol{t}}(\boldsymbol{x}) - \boldsymbol{t}]^{\mathsf{T}}\} - \boldsymbol{W}(\boldsymbol{H})$ is positive semi-definite, where $\boldsymbol{W}(\boldsymbol{H}) \triangleq \boldsymbol{H}\boldsymbol{G}^{-1}\boldsymbol{H}^{\mathsf{T}}$, and $\boldsymbol{G} \triangleq \boldsymbol{G}(\boldsymbol{H})$ is a $Q \times Q$ matrix whose elements are given by

$$[\boldsymbol{G}]_{a,b} = \frac{\mathbb{E}_{\boldsymbol{x},\boldsymbol{t}}\left\{ \begin{array}{c} \left(\sqrt{\frac{p(\boldsymbol{x},\boldsymbol{t}+\boldsymbol{h}_a)}{p(\boldsymbol{x},\boldsymbol{t})}} - \sqrt{\frac{p(\boldsymbol{x},\boldsymbol{t}-\boldsymbol{h}_a)}{p(\boldsymbol{x},\boldsymbol{t})}} \right) \\ \times \left(\sqrt{\frac{p(\boldsymbol{x},\boldsymbol{t}+\boldsymbol{h}_b)}{p(\boldsymbol{x},\boldsymbol{t})}} - \sqrt{\frac{p(\boldsymbol{x},\boldsymbol{t}-\boldsymbol{h}_b)}{p(\boldsymbol{x},\boldsymbol{t})}} \right) \end{array} \right\}}{\mathbb{E}_{\boldsymbol{x},\boldsymbol{t}}\left\{ \sqrt{\frac{p(\boldsymbol{x},\boldsymbol{t}+\boldsymbol{h}_a)}{p(\boldsymbol{x},\boldsymbol{t})}} \right\} \mathbb{E}_{\boldsymbol{x},\boldsymbol{t}}\left\{ \sqrt{\frac{p(\boldsymbol{x},\boldsymbol{t}+\boldsymbol{h}_b)}{p(\boldsymbol{x},\boldsymbol{t})}} \right\}}. \quad (6)$$

The tightest bound can then be obtained by maximizing $\boldsymbol{W}$ with respect to $\boldsymbol{H}$: $\textbf{WWB} \triangleq \sup_{\boldsymbol{H}} \boldsymbol{W}(\boldsymbol{H})$. In order to make the derivation of $\boldsymbol{G}$ tractable, we assume matrix $\boldsymbol{H}$ is diagonal, i.e., its column vectors $\boldsymbol{h}_q$ have their q-th component that is nonzero: $[\boldsymbol{h}_q]_i = h_q\, \delta_{i,q}$, where $\delta_{i,q}$ denotes the Krönecker delta. Note that the original version of the WWB from [2] also depends on auxiliary parameters $s_q \in \,]0\,, 1[$, $q = 1, \ldots, Q$, with respect to which the maximization of $\boldsymbol{W}$ has to be performed. However, it has been noticed that in a number of applications (see [6] for instance), the values $s_a = s_b = 1/2$ lead to the tightest bound. It is the case for the multiple change-point estimation problem as well, as it has been noticed in [8] after extensive simulations. For this reason, and to simplify the exposition, these values have been set into (6).

B. WWBs for the multiple change-point problem

In this section, we derive expressions of the WWBs for the multiple change-point estimation problem presented in Section II-A. Note that for this problem, the parameter space $\mathcal{T}$ is discrete and finite ($\mathcal{T} \subset \mathbb{N}^Q$), thus the set $\mathcal{H}$ and the set of lower bounds $\mathcal{W} \triangleq \{\boldsymbol{W}(\boldsymbol{H}) \,|\, \boldsymbol{H} \in \mathcal{H}\}$ are discrete and finite as well. Hence, as explained in [8], the supremum of $\mathcal{W}$ (that is the tightest WWB) can be obtained by finding the matrix $\boldsymbol{W}^\star$ associated with the minimum volume centered hyper-ellipsoid that covers all the centered hyper-ellipsoids associated with matrices $\boldsymbol{W} \in \mathcal{W}$. Such a task can actually be set as a convex optimization problem [11], which can be solved efficiently using an appropriate CVX toolbox.

Before doing so, we first need to obtain the expressions for the elements in matrix $\boldsymbol{G}$. By developing the numerator of (6), we obtain

$$[\boldsymbol{G}]_{a,b} = \frac{\begin{array}{r} \xi(\boldsymbol{h}_a, \boldsymbol{h}_b) + \xi(-\boldsymbol{h}_a, -\boldsymbol{h}_b) \\ -\, \xi(\boldsymbol{h}_a, -\boldsymbol{h}_b) - \xi(-\boldsymbol{h}_a, \boldsymbol{h}_b) \end{array}}{\xi(\boldsymbol{h}_a, \boldsymbol{0})\, \xi(\boldsymbol{h}_b, \boldsymbol{0})} \quad (7)$$

where

$$\begin{aligned} \xi(\boldsymbol{h}_a, \boldsymbol{h}_b) &\triangleq \mathbb{E}_{\boldsymbol{x},\boldsymbol{t}}\left\{ \frac{\sqrt{p(\boldsymbol{x}, \boldsymbol{t}+\boldsymbol{h}_a)\, p(\boldsymbol{x}, \boldsymbol{t}+\boldsymbol{h}_b)}}{p(\boldsymbol{x}, \boldsymbol{t})} \right\} \\ &= \sum_{\boldsymbol{t} \in \mathcal{T}} \nu(\boldsymbol{h}_a, \boldsymbol{h}_b, \boldsymbol{t})\, M(\boldsymbol{h}_a, \boldsymbol{h}_b, \boldsymbol{t}) \end{aligned} \quad (8)$$

with the following definitions of $\nu(\boldsymbol{h}_a, \boldsymbol{h}_b, \boldsymbol{t})$ and $M(\boldsymbol{h}_a, \boldsymbol{h}_b, \boldsymbol{t})$:

$$\nu(\boldsymbol{h}_a, \boldsymbol{h}_b, \boldsymbol{t}) \triangleq \sqrt{\pi(\boldsymbol{t}+\boldsymbol{h}_a)\, \pi(\boldsymbol{t}+\boldsymbol{h}_b)} \quad (9)$$

on the one hand, and

$$M(\boldsymbol{h}_a, \boldsymbol{h}_b, \boldsymbol{t}) \triangleq \int_{\Omega} \sqrt{p(\boldsymbol{x} \,|\, \boldsymbol{t}+\boldsymbol{h}_a)\, p(\boldsymbol{x} \,|\, \boldsymbol{t}+\boldsymbol{h}_b)}\, \mathrm{d}\boldsymbol{x} \quad (10)$$

on the other hand. Note that in (8), the discrete summation sign "$\sum_{\boldsymbol{t} \in \mathcal{T}}$" stands for the Q sums $\sum_{t_1} \ldots \sum_{t_Q}$.

We only derive the upper triangle terms from $\boldsymbol{G}$ (i.e., for $b \ge a$) and deduce the others by symmetry.

We use the following methodology to derive the elements in $\boldsymbol{G}$:

– derive $M(\boldsymbol{h}_a, \boldsymbol{h}_b, \boldsymbol{t})$;

– derive $\nu(\boldsymbol{h}_a, \boldsymbol{h}_b, \boldsymbol{t})$ for the various priors proposed in Section II-B;

– derive $\xi(\boldsymbol{h}_a, \boldsymbol{h}_b)$ (and deduce $[\boldsymbol{G}]_{a,b}$ when simplifications appear).

ISBN 978-90-827970-1-5

Expressions of $M(\boldsymbol{h}_a, \boldsymbol{h}_b, \boldsymbol{t})$ and $\nu(\boldsymbol{h}_a, \boldsymbol{h}_b, \boldsymbol{t})$ depend on whether we consider i) the diagonal terms ($b = a \triangleq d$), ii) the first super-diagonal terms ($b = a + 1$) or iii) the other terms in the upper triangle of $\boldsymbol{G}$ ($b > a + 1$).

1) Expressions of $M(\boldsymbol{h}_a, \boldsymbol{h}_b, \boldsymbol{t})$: These expressions do not depend on the chosen prior, and we derived them for the multiple change-point problem in [8]. Using the independence of the observations and the fact that only one component of $\boldsymbol{h}_a$ and $\boldsymbol{h}_b$ is nonzero (as explained in Section III-A), we find the following expressions for $M(\boldsymbol{h}_a, \boldsymbol{h}_b, \boldsymbol{t})$:

• if $b = a$ (diagonal terms of $\boldsymbol{G}$)

$$M(\boldsymbol{h}_a, \boldsymbol{h}'_a, \boldsymbol{t}) = \begin{cases} 1 & \text{if } h'_a = h_a \\ \rho_{a,a+1}^{|h_a|} & \text{if } h'_a = 0 \\ \rho_{a,a+1}^{2|h_a|} & \text{if } h'_a = -h_a \end{cases} \tag{11}$$

where, for $a = 1, \dots, Q$,

$$\rho_{a,a+1} \triangleq \int_{\Omega'} \sqrt{p_{\boldsymbol{\eta}_a}(x)\, p_{\boldsymbol{\eta}_{a+1}}(x)}\, \mathrm{d}x \tag{12}$$

denotes the Bhattacharyya distance between the distributions in the a-th and the $(a+1)$-th segments (Ω' denotes the single observation space, i.e., $\Omega' \times \ldots \times \Omega' = (\Omega')^N = \Omega$);

• if $b > a$ (upper triangle terms of $\boldsymbol{G}$), and $t_a + h_a \leq t_{a+1} + h_{a+1}$ (if $b = a + 1$)

$$M(\boldsymbol{h}_a, \boldsymbol{h}_b, \boldsymbol{t}) = M(\boldsymbol{h}_a, \boldsymbol{h}_b) = \rho_{a,a+1}^{|h_a|}\, \rho_{b,b+1}^{|h_b|} \tag{13}$$

• "overlap" between two consecutive test-points, that is, if $b = a + 1$, $h_a = |h_a| > 0$, $h_{a+1} = -|h_{a+1}| < 0$ and $t_a + h_a > t_{a+1} + h_{a+1}$ (with $a < Q$)

$$M(\boldsymbol{h}_a, \boldsymbol{h}_{a+1}, \boldsymbol{t}) = \rho_{a,a+1}^{|h_a|}\, \rho_{a+1,a+2}^{|h_{a+1}|}\, R_a^{t_{a+1}-|h_{a+1}|-t_a-|h_a|} \tag{14}$$

where, for $a = 1, \dots, Q - 1$,

$$R_a \triangleq \frac{\rho_{a,a+1}\, \rho_{a+1,a+2}}{\rho_{a,a+2}}. \tag{15}$$

Note that in this last case only, $M(\boldsymbol{h}_a, \boldsymbol{h}_b, \boldsymbol{t})$ actually depends on $\boldsymbol{t}$. In the following, in order to make these dependences or independences on $\boldsymbol{t}$ more visible, we denote $M(\boldsymbol{h}_a, \boldsymbol{h}_{a+1}, \boldsymbol{t})$ by $\tilde{M}(\boldsymbol{h}_a, \boldsymbol{h}_{a+1}, \boldsymbol{t})$ in the "overlap" case only, whereas we will denote $M(\boldsymbol{h}_a, \boldsymbol{h}_{a+1}, \boldsymbol{t})$ merely by $M(\boldsymbol{h}_a, \boldsymbol{h}_b)$ in all the other cases.

2) Expressions of $\nu(\boldsymbol{h}_a, \boldsymbol{h}_b, \boldsymbol{t})$ for a general prior: To expose the methodology, let us consider a generic prior distribution for the change-points that covers those introduced in Section II-B, i.e.,

$$\pi(\boldsymbol{t}) = \prod_{q=1}^{Q} \pi(t_q \,|\, t_{q-1}) \tag{16}$$

with $t_0 \triangleq 0$.

• For $b = a$ (diagonal terms of $\boldsymbol{G}$), we have $\nu(\boldsymbol{h}_a, \boldsymbol{h}_a, \boldsymbol{t}) = \pi(\boldsymbol{t} + \boldsymbol{h}_a)$ on the one hand, and on the other hand, for $a < Q$, we have

$$\nu(\boldsymbol{h}_a, -\boldsymbol{h}_a, \boldsymbol{t}) = \nu(2\boldsymbol{h}_a, \boldsymbol{0}, \boldsymbol{t}') = \left(\prod_{\substack{q=1\\ q\neq a,a+1}}^{Q} \pi(t'_q \,|\, t'_{q-1})\right) \sqrt{\pi(t'_a + 2h_a \,|\, t'_{a-1})\, \pi(t'_a \,|\, t'_{a-1})} \times \sqrt{\pi(t'_{a+1} \,|\, t'_a + 2h_a)\, \pi(t'_{a+1} \,|\, t'_a)} \tag{17}$$

and for $a = Q$, we have

$$\nu(\boldsymbol{h}_Q, -\boldsymbol{h}_Q, \boldsymbol{t}) = \nu(2\boldsymbol{h}_Q, \boldsymbol{0}, \boldsymbol{t}') = \left(\prod_{q=1}^{Q-1} \pi(t'_q \,|\, t'_{q-1})\right) \sqrt{\pi(t'_Q + 2h_Q \,|\, t'_{Q-1})\, \pi(t'_Q \,|\, t'_{Q-1})}, \tag{18}$$

after setting $\boldsymbol{t}' = \boldsymbol{t} - \boldsymbol{h}_a$ to obtain (17) and (18).

• For $b = a + 1$ (superdiagonal terms of $\boldsymbol{G}$), and $a < Q - 1$, we have

$$\nu(\boldsymbol{h}_a, \boldsymbol{h}_{a+1}, \boldsymbol{t}) = \left(\prod_{\substack{q=1\\ q\neq a,a+1,a+2}}^{Q} \pi(t_q \,|\, t_{q-1})\right) \sqrt{\pi(t_a + h_a \,|\, t_{a-1})\, \pi(t_a \,|\, t_{a-1})} \times \sqrt{\pi(t_{a+1} \,|\, t_a + h_a)\, \pi(t_{a+1} + h_{a+1} \,|\, t_a)} \times \sqrt{\pi(t_{a+2} \,|\, t_{a+1} + h_{a+1})\, \pi(t_{a+2} \,|\, t_{a+1})} \tag{19}$$

As in the previous case, if $a = Q - 1$, the last factor (with t_{a+2}) in (19) disappears.

• For $b > a + 1$ (other upper triangle terms of $\boldsymbol{G}$), finally, similar expression can be obtained. If $b < Q$, we have

$$\nu(\boldsymbol{h}_a, \boldsymbol{h}_b, \boldsymbol{t}) = \left(\prod_{\substack{q=1\\ q\neq a,a+1,b,b+1}}^{Q} \pi(t_q \,|\, t_{q-1})\right) \sqrt{\pi(t_a + h_a \,|\, t_{a-1})\, \pi(t_a \,|\, t_{a-1})} \times \sqrt{\pi(t_{a+1} \,|\, t_a + h_a)\, \pi(t_{a+1} \,|\, t_a)} \times \sqrt{\pi(t_b + h_b \,|\, t_{b-1})\, \pi(t_b \,|\, t_{b-1})} \times \sqrt{\pi(t_{b+1} \,|\, t_b + h_b)\, \pi(t_{b+1} \,|\, t_b)}, \tag{20}$$

and if $b = Q$, as in the previous cases, the last factor (with t_{b+1}) in (20) disappears.

3) Derivation of $\xi(\boldsymbol{h}_a, \boldsymbol{h}_b)$: For the derivation of $\xi(\boldsymbol{h}_a, \boldsymbol{h}_b)$, some attention has to be given to the bounds of the Q discrete sums in (8), since the support of the prior distribution is finite. The general form of the summation domain in (8) is

$$\mathcal{T}'_{-h_a,-h_b} \triangleq \mathcal{T} \cap (\mathcal{T} - \boldsymbol{h}_a) \cap (\mathcal{T} - \boldsymbol{h}_b) \tag{21}$$

where, with abuse of notations, we denote $\mathcal{T} - \boldsymbol{h} \triangleq \{\boldsymbol{t} - \boldsymbol{h} \,|\, \boldsymbol{t} \in \mathcal{T}\}$, and we recall that $\mathcal{T}$ is the support of the prior distribution, that can be deduced from the range of the indicator functions in its expression, cf. Section II-B.

• For $b = a$ (diagonal terms of $\boldsymbol{G}$), after plugging (11) into (8), we have, on the one hand,

$$\xi(\boldsymbol{h}_a, \boldsymbol{h}_a) = \sum_{\boldsymbol{t} \in \mathcal{T}'_{-h_a,0}} \pi(\boldsymbol{t} + \boldsymbol{h}_a), \tag{22}$$

that is generally lower than 1, and, on the other hand,

$$\begin{aligned}\xi(\boldsymbol{h}_a, -\boldsymbol{h}_a) = \xi(-\boldsymbol{h}_a, \boldsymbol{h}_a) &= \rho_{a,a+1}^{2|h_a|} \sum_{\boldsymbol{t}' \in \mathcal{T}'_{-2h_a,0}} \nu(2\boldsymbol{h}_a, \boldsymbol{0}, \boldsymbol{t}') \\ &= \rho_{a,a+1}^{2|h_a|} \sum_{t'_1} \pi(t'_1) \sum_{t'_2} \pi(t'_2 \,|\, t'_1) \ldots \\ &\quad \ldots \sum_{t'_a} \sqrt{\pi(t'_a + 2h_a \,|\, t'_{a-1})\, \pi(t'_a \,|\, t'_{a-1})}\end{aligned}$$

$$\sum_{t'_{a+1}} \sqrt{\pi(t'_{a+1}\,|\,t'_a+2h_a)\,\pi(t'_{a+1}\,|\,t'_a)}\dots$$
$$\dots\sum_{t'_{a+2}} \pi(t'_{a+2}\,|\,t'_{a+1})\dots\sum_{t'_Q}\pi(t'_Q\,|\,t'_{Q-1}). \quad (23)$$

Note that the $Q-(a+1)$ last sums in (23) equal 1. Consequently, after plugging (22) and (23) into (7), we obtain

$$\mathrm{num}([\boldsymbol{G}]_{a,a}) = \sum_{\boldsymbol{t}\in\mathcal{T}'_{-h_a,0}} \pi(\boldsymbol{t}+\boldsymbol{h}_a) + \sum_{\boldsymbol{t}\in\mathcal{T}'_{h_a,0}} \pi(\boldsymbol{t}-\boldsymbol{h}_a) - 2\rho_{a,a+1}^{2|h_a|}\sum_{\boldsymbol{t}'\in\mathcal{T}'_{-2h_a,0}}\nu(2\boldsymbol{h}_a,\boldsymbol{0},\boldsymbol{t}') \quad (24)$$

and

$$\mathrm{den}([\boldsymbol{G}]_{a,a}) = \rho_{a,a+1}^{2|h_a|}\left(\sum_{\boldsymbol{t}'\in\mathcal{T}'_{-h_a,0}}\nu(\boldsymbol{h}_a,\boldsymbol{0},\boldsymbol{t}')\right)^2 \quad (25)$$

in which $\mathrm{num}([\boldsymbol{G}]_{a,a})$ and $\mathrm{den}([\boldsymbol{G}]_{a,a})$ respectively denote the numerator and the denominator of $[\boldsymbol{G}]_{a,a}$ from (7).

• For $b=a+1$ (superdiagonal terms of $\boldsymbol{G}$), in cases with no overlap between the test points, we obtain expressions for $\xi(\boldsymbol{h}_a,\boldsymbol{h}_{a+1})$ similar to (23), using (13) and (19):

$$\xi(\boldsymbol{h}_a,\boldsymbol{h}_{a+1}) = \rho_{a,a+1}^{|h_a|}\,\rho_{a+1,a+2}^{|h_{a+1}|}\sum_{\boldsymbol{t}\in\mathcal{T}'_{-h_a,-h_{a+1}}}\nu(\boldsymbol{h}_a,\boldsymbol{h}_{a+1},\boldsymbol{t}). \quad (26)$$

If $h_a=|h_a|>0$ and $h_{a+1}=-|h_{a+1}|<0$, we have to split the $(a+1)$-st sum in (8) between the “overlap terms” for which $M(\boldsymbol{h}_a,\boldsymbol{h}_{a+1},\boldsymbol{t})=\tilde{M}(\boldsymbol{h}_a,\boldsymbol{h}_{a+1},\boldsymbol{t})$, for $t_{a+1}<t_a+|h_a|+|h_{a+1}|$, and the “non overlap terms” for which $M(\boldsymbol{h}_a,\boldsymbol{h}_{a+1},\boldsymbol{t})=M(\boldsymbol{h}_a,\boldsymbol{h}_{a+1})$, for $t_{a+1}\geq t_a+|h_a|+|h_{a+1}|$. This yields

$$\xi(|\boldsymbol{h}_a|,-|\boldsymbol{h}_{a+1}|) = \rho_{a,a+1}^{|h_a|}\,\rho_{a+1,a+2}^{|h_{a+1}|}(S_1+S_2) \quad (27)$$

in which

$$\begin{aligned} S_1 = &\sum_{t_1}\pi(t_1)\sum_{t_2}\pi(t_2\,|\,t_1)\dots \\ &\dots\sum_{t_a}\frac{\sqrt{\pi(t_a+h_a\,|\,t_{a-1})\,\pi(t_a\,|\,t_{a-1})}}{R_a^{t_a+|h_a|+|h_{a+1}|}} \\ &\sum_{t_{a+1}<t_a+|h_a|+|h_{a+1}|} R_a^{t_{a+1}}\sqrt{\pi(t_{a+1}\,|\,t_a+h_a)\,\pi(t_{a+1}+h_{a+1}\,|\,t_a)} \\ &\sum_{t_{a+2}}\sqrt{\pi(t_{a+2}\,|\,t_{a+1}+h_{a+1})\,\pi(t_{a+2}\,|\,t_{a+1})} \end{aligned} \quad (28)$$

and

$$\begin{aligned} S_2 = &\sum_{t_1}\pi(t_1)\sum_{t_2}\pi(t_2\,|\,t_1)\dots \\ &\dots\sum_{t_a}\sqrt{\pi(t_a+h_a\,|\,t_{a-1})\,\pi(t_a\,|\,t_{a-1})} \\ &\sum_{t_{a+1}\geq t_a+|h_a|+|h_{a+1}|}\sqrt{\pi(t_{a+1}\,|\,t_a+h_a)\,\pi(t_{a+1}+h_{a+1}\,|\,t_a)} \\ &\sum_{t_{a+2}}\sqrt{\pi(t_{a+2}\,|\,t_{a+1}+h_{a+1})\,\pi(t_{a+2}\,|\,t_{a+1})} \end{aligned} \quad (29)$$

after seeing the $Q-(a+2)$ last sums equal 1. This leads to

$$\begin{aligned} &\mathrm{num}([\boldsymbol{G}]_{a,a+1}) \\ &= \Bigg(\sum_{\boldsymbol{t}\in\mathcal{T}'_{-|h_a|,-|h_{a+1}|}}\nu(|\boldsymbol{h}_a|,|\boldsymbol{h}_{a+1}|,\boldsymbol{t}) + \sum_{\boldsymbol{t}\in\mathcal{T}'_{|h_a|,|h_{a+1}|}}\nu(-|\boldsymbol{h}_a|,-|\boldsymbol{h}_{a+1}|,\boldsymbol{t}) \\ &\quad - \sum_{\boldsymbol{t}\in\mathcal{T}'_{|h_a|,-|h_{a+1}|}}\nu(-|\boldsymbol{h}_a|,|\boldsymbol{h}_{a+1}|,\boldsymbol{t}) - S_1 - S_2\Bigg)\mathrm{sign}(h_a\,h_{a+1}) \end{aligned} \quad (30)$$

and

$$\mathrm{den}([\boldsymbol{G}]_{a,a+1}) = \left(\sum_{\boldsymbol{t}'\in\mathcal{T}'_{-h_a,0}}\nu(\boldsymbol{h}_a,\boldsymbol{0},\boldsymbol{t}')\right)\left(\sum_{\boldsymbol{t}'\in\mathcal{T}'_{-h_{a+1},0}}\nu(\boldsymbol{h}_{a+1},\boldsymbol{0},\boldsymbol{t}')\right) \quad (31)$$

• For $b>a+1$ (other upper triangle terms of $\boldsymbol{G}$), from (13) and (20), it can be shown that $\xi(\boldsymbol{h}_a,\boldsymbol{h}_b)=-\xi(\boldsymbol{h}_a,-\boldsymbol{h}_b)$ by using the variable change $t'_b=t_b-h_b$ in the latter. Consequently, we also have $\xi(-\boldsymbol{h}_a,-\boldsymbol{h}_b)=-\xi(-\boldsymbol{h}_a,\boldsymbol{h}_b)$. This implies, from (7), that, for $|b-a|>1$,

$$[\boldsymbol{G}]_{a,b}=0 \quad (32)$$

i.e., the matrix $\boldsymbol{G}$ has a tridiagonal structure. This was shown in [8] for the particular case of the URW prior.

C. Expressions of the WWB for the different proposed priors

1) Uniform on independent segments (UIS): For the UIS prior, it is worth noticing that the range of admissible test-points h_q is $[\![1-\lfloor\frac{N}{Q}\rfloor,\lfloor\frac{N}{Q}\rfloor-1]\!]$. Consequently, it is incompatible with the condition that there is at least one “overlap term” in S_1. Indeed, for $h_a>0$ and $h_{a+1}<0$, the upper bound of the sum w.r.t. t_a in (8) is $\tau_a-|h_a|$, while the lower bound w.r.t. t_{a+1} is $\tau_a+|h_{a+1}|+1$, which implies that the condition $t_a+|h_a|>t_{a+1}-|h_{a+1}|$ cannot be met, given the range of admissible values for h_a. Hence, for this prior distribution, we obtain that $[\boldsymbol{G}_{a,a+1}]=0$ as well, i.e., the matrix $\boldsymbol{G}$ is diagonal, whose diagonal terms are given by

$$[\boldsymbol{G}_{a,a}] = \frac{2\left(1-\frac{|h_a|}{\lfloor N/Q\rfloor}-\rho_{a,a+1}^{2|h_a|}\left(1-\frac{2|h_a|}{\lfloor N/Q\rfloor}\right)\right)}{\rho_{a,a+1}^{2|h_a|}\left(1-\frac{|h_a|}{\lfloor N/Q\rfloor}\right)^2}. \quad (33)$$

2) Uniform random walk (URW): The results for the URW prior are given in [8] in the special case $d=1$. Analogous expressions can be obtained using the following results:

$$\sum_{\boldsymbol{t}\in\mathcal{T}'_{-h_a,0}}\pi(\boldsymbol{t}\pm\boldsymbol{h}_a) = \sum_{\boldsymbol{t}'\in\mathcal{T}'_{-h_a,0}}\nu(\boldsymbol{h}_a,\boldsymbol{0},\boldsymbol{t}') = \begin{cases}\left(1-\frac{|h_a|}{\Delta}\right)^2 & \text{if } a<Q\\ 1-\frac{|h_Q|}{\Delta} & \text{if } a=Q\\ 0, & \text{otherwise,}\end{cases} \quad (34)$$

$$\sum_{\boldsymbol{t}\in\mathcal{T}'_{-h_a,-h_{a+1}}}\nu(\boldsymbol{h}_a,\boldsymbol{h}_{a+1},\boldsymbol{t}) = \begin{cases}\frac{(\Delta-|h_a|)(\Delta-|h_{a+1}|)}{\Delta^3}, & \text{if } a<Q\\ \frac{\Delta-|h_Q|}{\Delta^2}, & \text{if } a=Q\\ 0, & \text{otherwise,}\end{cases} \quad (35)$$

ISBN 978-90-827970-1-5

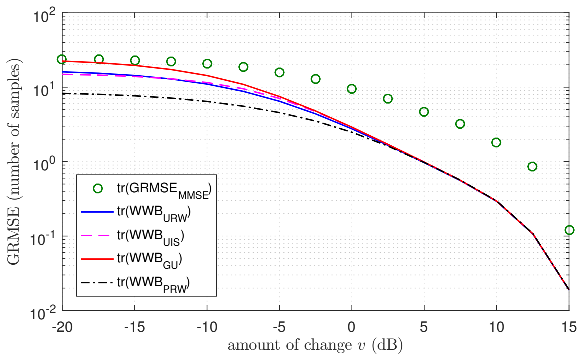

Figure 1. Simulation results in the case of $Q = 2$ unknown changes in the mean of $N = 80$ observations. Empirical GRMSE of the MMSE estimator of the change-points (green circles), and associated WWBs for the different priors introduced in Section II-B.

$$S_1 + S_2 = \left(\sum_{\boldsymbol{t}\in\mathcal{T}'_{-h_a,-h_{a+1}}} \nu(\boldsymbol{h}_a, \boldsymbol{h}_{a+1}, \boldsymbol{t})\right)\left(D - |h_a| - |h_{a+1}| + 1 - \frac{1 - R_a^{d-\min\{|h_a|,|h_{a+1}|\}}}{1 - R_a}\right). \tag{36}$$

It is worth noticing that, with these superdiagonal terms, the WWB directly depends on the distance between the successive change-points through parameters d and D.

3) Globally uniform distribution (GU): Considering the GU prior from (4), we obtain the following results:

$$\sum_{\boldsymbol{t}\in\mathcal{T}'_{-h_a,0}} \pi(\boldsymbol{t} \pm \boldsymbol{h}_a) = \sum_{\boldsymbol{t}'\in\mathcal{T}'_{-h_a,0}} \nu(\boldsymbol{h}_a, \boldsymbol{0}, \boldsymbol{t}') = \frac{\binom{N-1-|h_a|}{Q}}{\binom{N-1}{Q}}, \tag{37}$$

$$\sum_{\boldsymbol{t}\in\mathcal{T}'_{-h_a,-h_{a+1}}} \nu(\boldsymbol{h}_a, \boldsymbol{h}_{a+1}, \boldsymbol{t}) = \frac{\binom{N-1-|h_a|-|h_{a+1}|}{Q}}{\binom{N-1}{Q}}, \tag{38}$$

and

$$S_1 + S_2 = \frac{\binom{N-|h_a|-|h_{a+1}|}{Q}}{\binom{N-1}{Q}} + \sum_{t_1} \cdots \sum_{t_{a-1}} \sum_{t_a} \left(R_a^{-t_a-|h_a|-|h_{a+1}|} \sum_{t_{a+1}} \frac{\binom{N-1-t_{a+1}}{Q-(a+1)}}{\binom{N-1}{Q}} R_a^{t_{a+1}} \right) \tag{39}$$

where $t_q \in [\![t_{q-1}+1\,, N-Q+q-1]\!]$, $q = 1,\ldots,a-1$, $t_a \in [\![t_{a-1}+1\,, N-Q+a-|h_a|-1]\!]$, and $t_{a+1} \in [\![t_a + 1 + \max\{|h_a|, |h_{a+1}|\}\,, t_a + |h_a| + |h_{a+1}| - 1]\!]$.

4) Truncated Poisson random walk (PRW): Given its more complex expression, the PRW prior does not have any interesting closed form expression. However, it fits the kind of prior described by (16), thus the results from Section III-B apply. Since only discrete and finite sums appear in the expressions for the elements of matrix $\boldsymbol{G}$, the computation of the WWB with this prior is possible, although computationally more costly than with the aforementioned other prior distributions.

IV. Numerical results and concluding remarks

In this section, we present numerical results obtained in the case of changes in the mean of Gaussian observations, i.e., for $q = 1,\ldots,Q$ and $n = t_{q-1}+1,\ldots,t_q$, we have $x_n \sim \mathcal{N}(\mu_q, \sigma^2)$. Expressions of $\rho_{a,a+1}$ and R_a for this specific distribution have been derived in [8], and depend essentially on the so-called "amount of change" (also sometimes called signal-to-noise ratio, SNR), that we will denote by v. In the case of mean changes, the amount of change between the q-th and the $q+1$-th segment is defined by $v_{q,q+1} \triangleq (\mu_{q+1} - \mu_q)^2/\sigma^2$. In our simulations, we consider $Q = 2$, and we set $\sigma^2 = 1$, $\mu_1 = 1$. For a given amount of change v, we set $\mu_2 = \mu_1 + \sqrt{v\,\sigma^2}$, and $\mu_3 = \mu_1$. The evolution of the empirical global root mean square error (GRMSE) of the minimum mean square error (MMSE) estimator and the associated WWBs are displayed in Figure 1. The MMSE estimator is computed for the GU prior, and its squared error is averaged over 5000 Monte-Carlo runs. At each run, two new change-points t_1 and t_2 are randomly generated according to the GU prior. Asymptotically (in terms of amount of change), the WWBs tend to a common value regardless of the prior. In practice, this effect is visible for an amount of change v higher than 5 dB. Conversely, as the amount of change decreases, the WWBs differ progressively depending on the prior. It is worth mentioning that the chosen prior plays a significant role in the computation of the WWB. In the case of multiple change-point estimation, the PRW prior can be well-suited for some applications, but the computation of the WWB is very involved. On the contrary, the URW prior provides an interesting trade-off between computation complexity and the parameter space exploration.

References

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

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

[3] E. W. Barankin, "Locally best unbiased estimates," *Ann. Math. Stat.*, vol. 20, no. 4, pp. 477–501, Dec. 1949.

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

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

[6] D. T. Vu, A. Renaux, R. Boyer, and S. Marcos, "Some results on the Weiss-Weinstein bound for conditional and unconditional signal models in array processing," *Signal Process.*, vol. 95, no. 2, pp. 126–148, Feb. 2014.

[7] F. Xaver, C. F. Mecklenbrauker, P. Gerstoft, and G. Matz, "Weiss-Weinstein bounds for various priors," in *Proc. IEEE Workshop Sensor Array Multi-channel Process.*, Rio de Janeiro, Brazil, Jul. 2016, pp. 1–5.

[8] L. Bacharach, A. Renaux, M. N. El Korso, and E. Chaumette, "Weiss-Weinstein bound on multiple change-points estimation," *IEEE Trans. Signal Process.*, vol. 65, no. 10, pp. 2686–2700, May 2017.

[9] A. Cleynen and S. Robin, "Comparing change-point location in independent series," *Stat. Comput.*, vol. 26, no. 1-2, pp. 263–276, 2016.

[10] A. Mohammad-Djafari and O. Féron, "Bayesian approach to change points detection in time series," *Int. J. Imag. Syst. Tech.*, vol. 16, no. 5, pp. 215–221, Jan. 2006.

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