BAYESIAN-VARIATIONAL SCHEME TO AUTOMATICALLY SELECT THE REGULARIZATION PARAMETER IN THE ℓ_1-POTTS MODEL

Jordan Frecon[1]*, Nelly Pustelnik*[1]*, Nicolas Dobigeon*[2]*, Herwig Wendt*[2]*, Patrice Abry*[1]

[1] Univ Lyon, Ens de Lyon, Univ Claude Bernard, CNRS, Laboratoire de Physique, F-69342 Lyon, France `firstname.lastname@ens-lyon.fr`
[2] IRIT, CNRS, INP-ENSEEIHT, Toulouse, France `firstname.lastname@irit.fr`

ABSTRACT

This work focusses on the parameter estimation in the ℓ_1-Potts model, i.e., a variational approach involving a ℓ_1 dataterm and a TVℓ_0 penalization. Variational approaches based on total variation have gained considerable interest to solve piecewise constant denoising problems due to their deterministic setting and low computational cost. However, the estimation performance of the achieved solution strongly depend on the tuning of a regularization parameter. While recent works have tailored various hierarchical Bayesian procedure in order to additionally estimate the regularization parameter when the noise is Gaussian, less attention has been devoted to Laplacian noise. In this context and elaborating on a previous work, this contribution promotes a fast and parameter-free denoising procedure of piecewise constant signals corrupted by Laplacian noise. It relies on the minimization of a Bayesian-driven criteria whose similarities with the ℓ_1-Potts model permits to derive an efficient algorithm.

Index Terms— Piecewise constant denoising, Laplacian noise, regularization parameter, Potts model, hierarchical Bayesian model.

1. INTRODUCTION

The denoising of piecewise constant signals is of considerable interest in various applications [1, 2]. A large part of the literature relies on the assumption that the noise is white and Gaussian. However, in presence of outliers, it is more advisable to assume the noise Laplacian rather than Gaussian since the heavy-tail nature of the Laplacian distribution provides a more relevant modelisation of outliers [3, 4].

The denoising of piecewise constant signals corrupted by Laplace noise traced back to [5, 6]. Since then, it has rarely been envisaged in the Bayesian literature since, as opposed to Gaussian noise, posterior distributions are generally not expressible in terms of simple tractable functions of the observations. However, various methods have been developed around a generalized likelihood ratio test (see, e.g., [4] and references therein). In the variational context, there exists an efficient formulation based on the following ℓ_1-TVℓ_0 problem (usually referred to ℓ_1-Potts model) [7, 8]

$$\widehat{\boldsymbol{x}}_\lambda \in \underset{\boldsymbol{x}\in\mathbb{R}^N}{\operatorname{Argmin}} \|\boldsymbol{y}-\boldsymbol{x}\|_1 + \lambda\|L\boldsymbol{x}\|_0, \tag{1}$$

where $\boldsymbol{y} \in \mathbb{R}^N$ denotes the noisy data, $L \in \mathbb{R}^{(N-1)\times N}$ is the first difference operator, i.e., $(L\boldsymbol{x})_k = x_{k+1} - x_k$, and $\lambda \geq 0$ is a regularization parameter aiming at balancing the contribution of the data fidelity term and the regularization term. While (1) can be solved efficiently by means of dynamic programming [7, 8], the estimation performance of $\widehat{\boldsymbol{x}}_\lambda$ strongly depends on the choice of λ which is *a priori* unknown.

Related works. In the context of Gaussian noise and, thus, of the ℓ_2-TVℓ_q counterpart of (1) with $q = \{0, 1\}$, interesting ideas to select λ rely on a hierarchical Bayesian structure between $\boldsymbol{y}$, $\boldsymbol{x}$ and λ. For instance the selection of λ in the ℓ_2-TVℓ_1 problem can be solved by assuming a certain prior for λ and maximizing either the conditional distribution of $(\boldsymbol{x}, \lambda)$ given $\boldsymbol{y}$ [9, 10] or a marginalised posterior in order to remove λ from the model [11, 10]. In the context of ℓ_2-TVℓ_0, a similar procedure relying on a reparametrization of $\boldsymbol{x}$ is derived in [12] and its efficiency is proved through numerous experiments. All these approaches share the same idea that amounts in considering a joint estimation of $\boldsymbol{x}$ and λ, so requiring an additional penalization term designed from hierarchical Bayesian arguments.

Contributions and outline. Inspired from our previous contribution dedicated to Gaussian noise [12], we derive in this work a parameter-free estimation procedure suited to Laplacian piecewise contant denoising. It consists in solving

$$(\widehat{\boldsymbol{x}}, \widehat{\sigma}, \widehat{\lambda}) \in \underset{\boldsymbol{x}\in\mathbb{R}^N, \sigma>0, \lambda>0}{\operatorname{Argmin}} \frac{1}{\sigma}\|\boldsymbol{y}-\boldsymbol{x}\|_1 + \frac{\lambda}{\sigma}\|L\boldsymbol{x}\|_0 + \phi(\sigma, \lambda) \tag{2}$$

and having beforehand defined ϕ. To do so, a reparametrization of $\boldsymbol{x}$ is necessary and is recalled in Section 2. Then, the derivation of ϕ using hierarchical Bayesian arguments is described in Section 3. We design in Section 4 an algorithmic procedure providing an approximate solution for (2) which benefits from efficient dynamic programming techniques solving (1). Estimation performance are then discussed in Section 5 and conclusions are drawn in Section 6.

2. PROBLEM REPARAMETRIZATION

2.1. Problem

The underlying problem consists in estimating a piecewise constant signal $\overline{\boldsymbol{x}} \in \mathbb{R}^N$ from the noisy observations $\boldsymbol{y} = \overline{\boldsymbol{x}} + \boldsymbol{\epsilon}$. The noise samples $(\boldsymbol{\epsilon}_i)_{1\leq i\leq N}$ are assumed to be independent and identically distributed (i.i.d.) zero mean Laplace variables with common but unknown scale parameter σ, i.e., $\boldsymbol{\epsilon}|\sigma \sim \text{Laplace}\,(\boldsymbol{0}, \sigma\boldsymbol{I}_N)$.

2.2. Reparametrization

Following [13, 14], any candidate solution $\boldsymbol{x}$ can be parametrized by the indicator vector $\boldsymbol{r} \in \{0, 1\}^N$ of its change-points and the vector of values taken between each change-point.

The indicator vector $\boldsymbol{r} = \big(r_i\big)_{1\leq i\leq N}$ is introduced as follows

$$r_i = \begin{cases} 1, & \text{if there is a change-point at time instant } i, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

By convention, $r_i = 1$ indicates that x_i is the last sample belonging to the current segment, and thus that x_{i+1} belongs to the next segment. Moreover, stating $r_N = 1$ ensures that $K = \sum_{i=1}^{N} r_i$.

For each $k \in \{1, \dots, K\}$, the set $\mathcal{R}_k \subset \{1, \dots, N\}$ is used to represent the set of time indices associated to the k-th segment. Therefore, $\mathcal{R}_k \cap \mathcal{R}_{k'} = \{\emptyset\}$ for $k \neq k'$ and $\cup_{k=1}^{K} \mathcal{R}_k = \{1, \cdots, N\}$. Hereafter, the notation $K_{\boldsymbol{r}}$ will be adopted to emphasize the dependence of the number K of segments on the indicator vector $\boldsymbol{r}$, i.e., $K = \|\boldsymbol{r}\|_0$.

The values taken on each segment of $\boldsymbol{x}$ can be encoded by introducing the vector $\boldsymbol{\mu} = (\mu_k)_{1 \leq k \leq K_{\boldsymbol{r}}}$ such that

$$(\forall k \in \{1, \dots, K_{\boldsymbol{r}}\})(\forall i \in \mathcal{R}_k) \quad x_i = \mu_k. \tag{4}$$

The parametrization leads to this first result whose derivation is a direct result from arguments described in [12].

Proposition 2.1. *Let $\boldsymbol{y} \in \mathbb{R}^N$ and $\phi \colon \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$. Problem* (2) *is equivalent to*

$$\underset{\substack{\boldsymbol{\theta}=\{\boldsymbol{r},\boldsymbol{\mu}\} \in \{0,1\}^N \times \mathbb{R}^{K_{\boldsymbol{r}}} \\ \lambda > 0, \sigma > 0}}{\text{minimize}} \left\{ \frac{1}{\sigma} \sum_{k=1}^{K_{\boldsymbol{r}}} \sum_{i \in \mathcal{R}_k} (y_i - \mu_k)^2 + \frac{\lambda}{\sigma}(K_{\boldsymbol{r}} - 1) + \phi(\sigma, \lambda) \right\}. \tag{5}$$

3. BAYESIAN DRIVEN DERIVATION OF ϕ

It is well established that a minimization problem involving a data term and a regularization terms can be related to the maximization of the posterior distribution (see e.g., [15]). A similar idea is used here when posterior distribution is obtained with a hierarchical Bayesian formulation in order to derive a connection with (5).

3.1. Hierarchical Bayesian model

First, according to the noise degradation model, the joint likelihood function of the observations $\boldsymbol{y}$ given the piecewise constant model $\{\boldsymbol{r}, \boldsymbol{\mu}\}$ and the scale parameter σ follows a Laplacian distribution, i.e.,

$$f(\boldsymbol{y}|\boldsymbol{r}, \boldsymbol{\mu}, \sigma) = \prod_{k=1}^{K_{\boldsymbol{r}}} \prod_{i \in \mathcal{R}_k} \frac{1}{\sigma} \exp\Big(- \frac{|\mu_k - y_i|}{\sigma} \Big). \tag{6}$$

To derive the posterior distribution, prior distributions over $\boldsymbol{r}$, $\boldsymbol{\mu}$, and σ have to be specified. In the literature, it is often encountered to assume that $(r_i)_{1 \leq i \leq N}$ are independent and identically distributed (i.i.d.) according to a Bernoulli distribution with hyper-parameter p [16, 17, 18]. The parameter p quantifies the prior probability of occurrence of a change, independently of the location:

$$f(\boldsymbol{r}|p) = \prod_{i=1}^{N-1} p^{r_i} (1-p)^{1-r_i}.$$

From a hierarchical Bayesian perspective, a natural choice for the posterior distribution of segment amplitudes $(\mu_k)_{1 \leq k \leq K_{\boldsymbol{r}}}$ would consist in electing independent conjugate Laplace prior distributions. However, such posterior distributions cannot be expressible in terms of simple tractable functions. To alleviate this problem, we choose a non-informative prior such as a uniform distribution between some $\mu_{\min}$ and $\mu_{\max}$, i.e.,

$$f(\boldsymbol{\mu}|\boldsymbol{r}) = \prod_{k=1}^{K_{\boldsymbol{r}}} (\mu_{\max} - \mu_{\min})^{-1} \imath_{[\mu_{\min}, \mu_{\max}]}(\mu_k). \tag{7}$$

In particular, if we choose $\mu_{\min} < \min(\boldsymbol{y})$ and $\mu_{\max} > \max(\boldsymbol{y})$, then (7) recasts into

$$f(\boldsymbol{\mu}|\boldsymbol{r}) = (\mu_{\max} - \mu_{\min})^{-K_{\boldsymbol{r}}}. \tag{8}$$

Following usual prior choices in hierarchical approaches [14], a scale-invariant non-informative Jeffreys prior is assigned on the scale parameter σ in order to account for the absence of prior knowledge on σ, i.e.,

$$f(\sigma) \propto \frac{1}{\sigma}, \tag{9}$$

while a conjugate Beta distribution $\mathcal{B}(\alpha_0, \alpha_1)$ is assigned to the unknown hyper-parameter p:

$$f(p) = \frac{\Gamma(\alpha_0 + \alpha_1)}{\Gamma(\alpha_0)\,\Gamma(\alpha_1)} p^{\alpha_1 - 1} (1-p)^{\alpha_0 - 1}. \tag{10}$$

Assuming the parameters $\boldsymbol{r}$, $\boldsymbol{\mu}$ and σ to be a priori independent, we can derive the following joint posterior distribution

$$f(\boldsymbol{\Theta}|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{r}, \boldsymbol{\mu}, \sigma)\, f(\boldsymbol{\mu}|\boldsymbol{r})\, f(\boldsymbol{r}|p) f(p) f(\sigma) \tag{11}$$

with $\boldsymbol{\Theta} = \{\boldsymbol{r}, \boldsymbol{\mu}, \sigma, p\}$. Then, minimizing the minus log joint posterior distribution (11) leads to:

$$\begin{aligned} \underset{\boldsymbol{\Theta}=\{\boldsymbol{r},\boldsymbol{\mu},\sigma^2,p\}}{\text{minimize}} \; & \frac{1}{\sigma} \sum_{k=1}^{K_{\boldsymbol{r}}} \sum_{i \in \mathcal{R}_k} |y_i - \mu_k| \\ & + (K_{\boldsymbol{r}} - 1) \left(\log\left(\frac{1-p}{p} \right) + \log(\mu_{\max} - \mu_{\min}) \right) \\ & + N \log(2\sigma) - (N-1)\log(1-p) + \log\sigma \\ & - (\alpha_1 - 1)\log p - (\alpha_0 - 1)\log(1-p) \\ & + \log(\mu_{\max} - \mu_{\min}). \end{aligned} \tag{12}$$

3.2. Problems equivalence and derivation of ϕ

The core idea to derive ϕ consists in establishing the similitudes between problems (5) and (12).

Proposition 3.1. *The minimizations problems* (5) *and* (12) *lead to a similar solution for the following parametrization of λ and choice of ϕ :*

$$\frac{\lambda}{\sigma} = \left(\log\left(\frac{1-p}{p} \right) + \log(\mu_{\max} - \mu_{\min}) \right) \tag{13}$$

and

$$\begin{aligned} \phi(\sigma, \lambda) = & N \log(2\sigma) + \log(\sigma) \\ & - \frac{\lambda}{\sigma}(N + \alpha_0 - 2) + (N + \alpha_0 - 1)\log(\mu_{\max} - \mu_{\min}) \\ & + (N + \alpha_0 + \alpha_1 - 3) \log\left(1 + \exp\left(\frac{\lambda}{\sigma} - \log(\mu_{\max} - \mu_{\min}) \right) \right). \end{aligned} \tag{14}$$

The principle of the proof consists in observing that both (5) and (12) contains the same data fidelity term, a term proportional to $(K_{\boldsymbol{r}} - 1)$ which lead to Eq. (13) by identification and an additional term independent of $\boldsymbol{r}$, namely

$$\begin{aligned} \phi(\sigma, p) = & N \log(2\sigma) - (N-1)\log(1-p) + \log\sigma \\ & - (\alpha_1 - 1)\log p - (\alpha_0 - 1)\log(1-p) \\ & + \log(\mu_{\max} - \mu_{\min}). \end{aligned} \tag{15}$$

Finally, by inverting the relation (13), ϕ can be parametrized in terms of (σ, λ) such as in (14).

Algorithm 1 Bayesian driven resolution of the TVℓ_1 problem

Input: Observed signal $\boldsymbol{y} \in \mathbb{R}^N$.
The predefined set of regularization parameters Λ.
Hyperparameters $\Phi = \{\alpha_0, \alpha_1, \mu_{\max} - \mu_{\min}\}$.

Iterations:

1: **for** $\lambda \in \Lambda$ **do**
2: Compute $\widehat{\boldsymbol{x}}_\lambda = \arg\min_{\boldsymbol{x}\in\mathbb{R}^N} \|\boldsymbol{y} - \boldsymbol{x}\|_1 + \lambda\|L\boldsymbol{x}\|_0$.
3: Compute $\widehat{\sigma}_\lambda = \|\boldsymbol{y} - \widehat{\boldsymbol{x}}_\lambda\|_2/\sqrt{N-1}$.
4: **end for**

Output: Solution $\{\widehat{\boldsymbol{x}}_{\widehat{\lambda}}, \widehat{\lambda}, \widehat{\sigma}_{\widehat{\lambda}}\}$ with $\widehat{\lambda} = \arg\min_{\lambda\in\Lambda} F(\widehat{x}_\lambda, \lambda, \widehat{\sigma}_\lambda)$

4. ALGORITHMIC SOLUTION

Now that we have derived the explicit expression (14) of ϕ, we aim at efficiently solving the nonconvex optimization problem (5) in a deterministic setting. Following [12], we propose to estimate λ on a predefined grid Λ and to solve $(\forall \lambda \in \Lambda)$

$$(\widehat{\boldsymbol{x}}_\lambda, \widehat{\sigma}_\lambda) \in \underset{\boldsymbol{x}\in\mathbb{R}^N, \sigma\geq 0}{\operatorname{Argmin}} \underbrace{\frac{1}{\sigma}\|\boldsymbol{y} - \boldsymbol{x}\|_1 + \frac{\lambda}{\sigma}\|L\boldsymbol{x}\|_0 + \phi(\sigma, \lambda)}_{F(\boldsymbol{x},\lambda,\sigma)}, \quad (16)$$

which we approximate by

$$(\forall \lambda \in \Lambda) \quad \begin{cases} \widehat{\boldsymbol{x}}_\lambda & = \arg\min_{\boldsymbol{x}\in\mathbb{R}^N} \|\boldsymbol{y} - \boldsymbol{x}\|_1 + \lambda\|L\boldsymbol{x}\|_0, \\ \widehat{\sigma}_\lambda & = \|\boldsymbol{y} - \widehat{\boldsymbol{x}}_\lambda\|_2/\sqrt{N-1}, \end{cases} \quad (17)$$

and then selecting the triplet $\{\widehat{\boldsymbol{x}}_{\widehat{\lambda}}, \widehat{\lambda}, \widehat{\sigma}_{\widehat{\lambda}}\}$ such that

$$\widehat{\lambda} = \arg\min_{\lambda\in\Lambda} F(\widehat{\boldsymbol{x}}_\lambda, \lambda, \widehat{\sigma}_\lambda). \quad (18)$$

The corresponding algorithm steps are reported in Algorithm 1. This approach permits to use the dynamic programming algorithm *Pottslab* developed in [7, 8] to solve the problem (1) for any $\lambda \in \Lambda$.

5. ESTIMATION PERFORMANCE

5.1. Experimental setting

Data $\boldsymbol{y}$ are synthesized in two steps. First, the change-point locations of $\overline{\boldsymbol{x}}$ are drawn i.i.d. given the change-point probability $\overline{p}$, and then the value taken on each segment is uniformly drawn between a minimal value $\overline{x}_{\min}$ and a maximal value $\overline{x}_{\max}$ also given beforehand. Second, we generate $\boldsymbol{\epsilon} \sim \text{Laplace}\,(\boldsymbol{0}, \sigma\boldsymbol{I}_N)$ and form $\boldsymbol{y} = \overline{\boldsymbol{x}} + \boldsymbol{\epsilon}$. Therefore the mean amplitude between successive segments is about $(\overline{x}_{\max} - \overline{x}_{\min})/3$ and its comparison w.r.t. noise power, that is the amplitude-to-noise-ratio, reads $\text{ANR} = (\overline{x}_{\max} - \overline{x}_{\min})/(3\sigma)$.

The hyperparameters are chosen as follow. A non-informative prior is used for the prior probability p by setting $\alpha_0 = \alpha_1 = 1$ so that the Beta distribution in (10) reduces to a uniform distribution over $(0, 1)$. In addition, $\mu_{\max}$ and $\mu_{\min}$ are chosen such as $\mu_{\min} < \min \boldsymbol{y}$, $\mu_{\max} > \max \boldsymbol{y}$ and $\mu_{\max} - \mu_{\min} = 10^4$. This choice is further discussed in Section 5.3.

In our experiments, the set Λ of discrete values for λ has been composed of 500 values equally spaced, in a $\log_{10}$-scale, between 10^{-5} and 10^5.

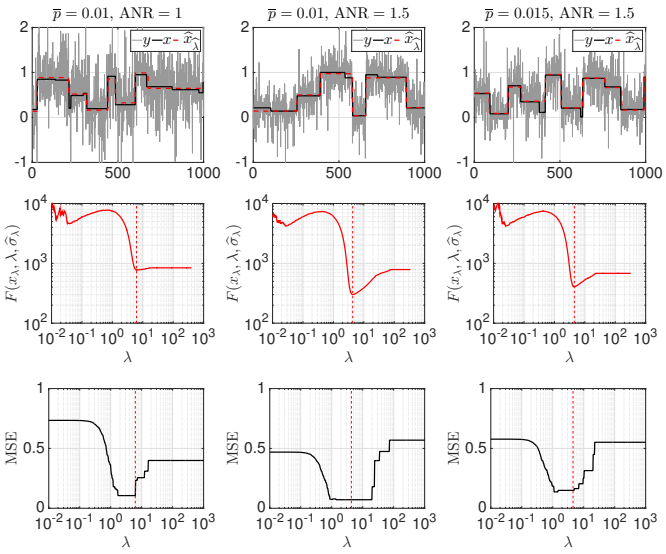

Fig. 1. **Illustration of the proposed method.** Three configurations are examined depending on $\overline{p}$ and ANR. The proposed criterion $F(\widehat{x}_\lambda, \lambda, \widehat{\sigma}_\lambda)$ is displayed in red in the second row as function of λ while the relative MSE between $\widehat{\boldsymbol{x}}_\lambda$ and $\overline{\boldsymbol{x}}$ is displayed in the third row. The estimate $\widehat{\lambda}$ is indicated in a vertical dashed line and the corresponding solution $\widehat{\boldsymbol{x}}_{\widehat{\lambda}}$ is reported in red in the first row.

5.2. Illustration of the automatic selection of λ

Three different observations $\boldsymbol{y}$ (grey) are represented in each columns of Fig.1 as representatives of different configurations of change-point probability p and scale parameter σ. The truth $\overline{\boldsymbol{x}}$ that we aim at recovering is display in black.

The proposed criterion $F(\widehat{x}_\lambda, \lambda, \widehat{\sigma}_\lambda)$ is displayed in the second row in solid red as function of λ. The position of its minimum $\widehat{\lambda}$ (see (18)) is indicated by a vertical dashed red line. The corresponding solution $\widehat{\boldsymbol{x}}_{\widehat{\lambda}}$ is reported in dashed red lines in the first row and appears as a visually good estimate of $\overline{\boldsymbol{x}}$.

Performance are further quantified in terms of relative mean squared error (MSE) between $\widehat{\boldsymbol{x}}_\lambda$ and $\overline{\boldsymbol{x}}$ as function of λ in the third row. For $p = 0.01$ and $\text{ANR} = 1.5$ (see Fig.1, middle), the method automatically selects one $\widehat{\lambda}$ such as the solution benefits from a lower MSE than for any other $\lambda \in \Lambda$. However, when the ANR decreases (see Fig.1, left) or $\overline{p}$ increases (see Fig.1, right), the solution may not be optimal but still maintains very good estimation performance close to the minimum MSE.

5.3. Quantification of estimation performance

In this section, we further examine the estimation performance with respect to (w.r.t.) the dynamic $\overline{x}_{\max} - \overline{x}_{\min} \in \{0.1, 1, 10\}$, the change-point probability $\overline{p} \in \{0.005, 0.01, 0.015\}$ and the ANR. The regularization parameter selected by the proposed method ($\widehat{\lambda}$, red) is compared against the similar method developed in [12] for Gaussian noise ($\widehat{\lambda}^{(\ell_2)}$, magenta). In addition, we also report the oracle estimate (Λ_{MSE}, dashed white) for which $\widehat{\boldsymbol{x}}_{\lambda\in\Lambda_{\text{MSE}}}$ yields the lowest MSE and whose range is delimited by two dashed white lines. Results presented here are averaged over 25 realizations.

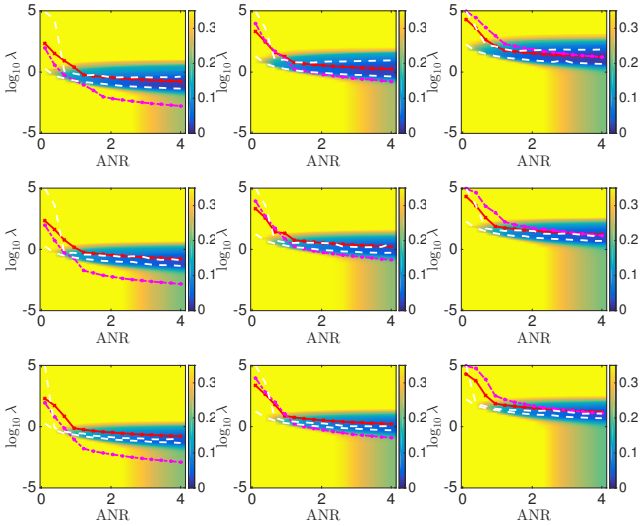

Fig. 2. **Estimation performance ($\widehat{\lambda}$ vs. ANR).** From top to bottom: $\overline{p} = 0.005$, 0.010 and 0.015. Different dynamics are examined from left to right: $\overline{x}_{\max} - \overline{x}_{\min} = 0.1$, 1, and 10. The mean proposed estimate $\widehat{\lambda}$ is displayed in red as a function of the ANR and is compared against $\widehat{\lambda}^{(\ell_2)}$ (mixed magenta) and the MSE oracle estimator Λ_{MSE} whose range is delimited by dashed white lines. Overall, the proposed solution (red line) remains between the dashed white lines, thus showing that $\widehat{\boldsymbol{x}}_\lambda$ for $\lambda = \widehat{\lambda}$ achieves the best performance in terms of relative MSE than for any other λ.

Behavior of $\widehat{\lambda}$. Each plot in Fig.2 illustrates how $\widehat{\lambda}$, $\widehat{\lambda}^{(\ell_2)}$ and Λ_{MSE} vary w.r.t. the ANR. For comparison purposes, the relative MSE is also superimposed. Overall, the proposed solution (red line) remains between the dashed white lines, thus showing that $\widehat{\boldsymbol{x}}_\lambda$ for $\lambda = \widehat{\lambda}$ achieves the best performance in terms of relative MSE than for any other λ. However, we observe that the performance deteriorate as $\overline{p}$ increases (see Fig.2 from top to bottom) as the estimation problem is more difficult when an higher number of segments need to be detected. This is quantified by the shrinking of the oracle range Λ_{MSE}.

Further investigations show that $\widehat{\lambda}$ scales properly when the dynamic varies (see translations of red and white lines in Fig.2 from left to right) whereas $\widehat{\lambda}^{(\ell_2)}$ does not. This means that if two observations are identical by a scale factor, then the proposed method outputs identical solutions by a scale factor.

Comparison of MSE. In addition, estimation performance are compared in terms of relative MSE between $\widehat{\boldsymbol{x}}_\lambda$ and $\overline{\boldsymbol{x}}$ for $\lambda = \widehat{\lambda}$ (red), $\widehat{\lambda}^{(\ell_2)}$ (magenta) and any $\lambda \in \Lambda_{\text{MSE}}$ (black) in Fig.3. Results illustrate that taking into account the Laplacian nature of the noise (red) rather than supposing it Gaussian (magenta) permits to systematically lower relative MSE. In addition, both methods exhibit similar performance for ANR large enough. Overall, the proposed method provides MSE performance as close to the oracle estimate as $\overline{p}$ is small and ANR is large.

Impact of hyperparameter $\mu_{\max} - \mu_{\min}$. For the same configurations as those presented in Fig. 1, the results displayed in Fig. 4 show that the tuning of $\mu_{\max} - \mu_{\min}$ does not require a complicate procedure as the estimation performance are the same for $\mu_{\max} - \mu_{\min}$ sufficiently large with respect to the amplitude dynamic of $\boldsymbol{y}$.

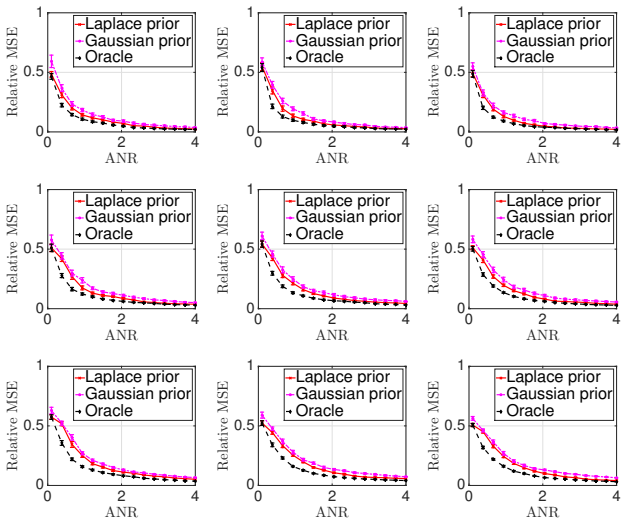

Fig. 3. **Estimation performance (MSE vs. ANR).** From top to bottom: $\overline{p} = 0.005$, 0.010 and 0.015. Different dynamics are examined from left to right: $\overline{x}_{\max} - \overline{x}_{\min} = 0.1$, 1, and 10. For each configuration, relative MSE $\|\widehat{\boldsymbol{x}}_\lambda - \overline{\boldsymbol{x}}\|/\|\overline{\boldsymbol{x}}\|$ are compared for $\lambda = \widehat{\lambda}$ (red), $\widehat{\lambda}^{(\ell_2)}$ (magenta) and any $\lambda \in \Lambda_{\text{MSE}}$ (black). Results show that including the knowledge that the noise is Laplacian permits to yield lower MSE.

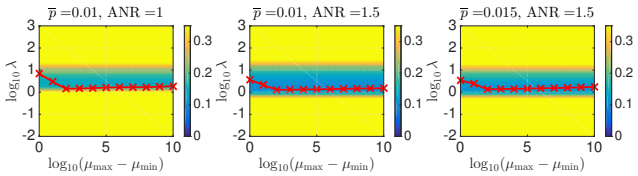

Fig. 4. **Impact of hyperparameters (MSE vs. $\mu_{\max} - \mu_{\min}$).** For each configuration, $\widehat{\lambda}$ (red) is plotted as a function of the hyperameter value $\mu_{\max} - \mu_{\min}$. Results show satisfactory and identical performances for any choice such as $\mu_{\max} - \mu_{\min} \geq 10^3$.

Computational cost. In the experiments presented here, simulations took around 40 seconds for $|\Lambda| = 500$ and $N = 10^3$.

6. CONCLUSION

Elaborating on previous work [12], the present contribution promotes the use of a bayesian-driven criteria to estimate the regularization parameter inherent to the ℓ_1-TVℓ_0 minimization problem. The criteria is derived by establishing the equivalence between the variational problem and a hierarchical Bayesian formulation of the change-point detection problem. The equivalence also permits to design an efficient algorithmic procedure which benefits from low computational costs. The good performance of the procedure were evaluated in terms of MSE and compared to oracle estimates.

7. REFERENCES

[1] M. Basseville and I.V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.

[2] M.A. Little and N.S. Jones, "Generalized methods and solvers for noise removal from piecewise constant signals. I. Background theory," *Proc. R. Soc. A*, vol. 467, pp. 3088–3114, 2011.

[3] M. Nikolova, "A variational approach to remove outliers and impulse noise," *J. Math. Imag. Vis.*, vol. 20, no. 1-2, pp. 99–120, 2004.

[4] A. Mincholé, L. Sörnmo, and P. Laguna, "Detection of body position changes from the ECG using a Laplacian noise model," *Biomed. Signal Process. Contr.*, vol. 14, pp. 189–196, 2014.

[5] R. J. Marks, G. L. Wise, D. G. Haldeman, and J. L. Whited, "Detection in Laplace noise," *IEEE Trans. on Aerospace and Electronic Systems*, vol. AES-14, no. 6, pp. 866–872, Nov 1978.

[6] M. Wu and W. J. Fitzgerald, "Analytical approach to change-point detection in Laplacian noise," *IEEE Proc. Vision, Image and Signal Processing*, vol. 142, no. 3, pp. 174–180, Jun 1995.

[7] M. Storath, A. Weinmann, and L. Demaret, "Jump-sparse and sparse recovery using Potts functionals," *IEEE Trans. Signal Process.*, vol. 62, no. 14, pp. 3654–3666, July 2014.

[8] M. Storath, A. Weinmann, and M. Unser, "Exact algorithms for l^1-TV regularization of real-valued or circle-valued signals," *J. Sci. Comput.*, vol. 38, no. 1, pp. 614–630, 2016.

[9] L. Chaari, J.-C. Pesquet, J.-Y. Tourneret, and P. Ciuciu, "Parameter estimation for hybrid wavelet-total variation regularization," in *Proc. IEEE Workshop Stat. Sign. Proc.*, Nice, France, June, 28-30 2011.

[10] M. Pereyra, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Maximum-a-posteriori estimation with unknown regularisation parameters," in *Proc. Eur. Sig. Proc. Conference*, Nice, France, Aug 2015, pp. 230–234.

[11] J. P. Oliveira, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Adaptive total variation image deblurring: a majorization–minimization approach," *Signal Process.*, vol. 89, no. 9, pp. 1683–1693, 2009.

[12] J. Frecon, N. Pustelnik, N. Dobigeon, H. Wendt, and P. Abry, "Bayesian selection for the regularization parameter in TVℓ_0 denoising problems," arXiv preprint arXiv:1608.07739, 2016.

[13] M. Lavielle, "Optimal segmentation of random processes," *IEEE Trans. Signal Process.*, vol. 46, no. 5, pp. 1365–1373, 1998.

[14] Nicolas Dobigeon, Jean-Yves Tourneret, and Manuel Davy, "Joint segmentation of piecewise constant autoregressive processes by using a hierarchical model and a Bayesian sampling approach," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1251–1263, Apr. 2007.

[15] N. Pustelnik, A. Benazza-Benhayia, Y. Zheng, and J.-C. Pesquet, "Wavelet-based image deconvolution and reconstruction," *Wiley Encyclopedia of EEE*, 2016.

[16] M. Lavielle and E. Lebarbier, "An application of MCMC methods for the multiple change-points problem," *Signal Process.*, vol. 81, no. 1, pp. 39–53, Jan. 2001.

[17] E. Punskaya, C. Andrieu, A. Doucet, and W. Fitzgerald, "Bayesian curve fitting using MCMC with applications to signal segmentation," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 747–758, Mar. 2002.

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