

Sélection du paramètre de régularisation dans le problème l2-Potts

Jordan Frécon, Nelly Pustelnik, Nicolas Dobigeon, Herwig Wendt, Patrice Abry

▶ **To cite this version:**

Jordan Frécon, Nelly Pustelnik, Nicolas Dobigeon, Herwig Wendt, Patrice Abry. Sélection du paramètre de régularisation dans le problème l2-Potts. 26eme Colloque GRETSI sur le Traitement du Signal et des Images (GRETSI 2017), Sep 2017, Juan-les-Pins, France. pp.1-4. hal-03658149

HAL Id: hal-03658149
https://hal.science/hal-03658149

Submitted on 3 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Open Archive Toulouse Archive Ouverte

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible

This is an author's version published in:
http://oatao.univ-toulouse.fr/19126

To cite this version: Frécon, Jordan and Pustelnik, Nelly and Dobigeon, Nicolas and Wendt, Herwig and Abry, Patrice *Sélection du paramètre de régularisation dans le problème l2-Potts.* (2017) In: 26eme Colloque GRETSI sur le Traitement du Signal et des Images (GRETSI 2017), 5 September 2017 - 8 September 2017 (Juan-les-Pins, France).

Any correspondence concerning this service should be sent to the repository administrator: tech-oatao@listes-diff.inp-toulouse.fr

Sélection du paramètre de régularisation dans le problème ℓ_2-Potts

Jordan FRÉCON[1], Nelly PUSTELNIK[2], Nicolas DOBIGEON[3], Herwig WENDT[3], Patrice ABRY[2]

[1]Computational Statistics and Machine Learning, Istituto Italiano di Tecnologia, Genoa, Italy

[2]Univ Lyon, Ens de Lyon, Univ Claude Bernard, CNRS, Laboratoire de Physique, F-69342 Lyon, France

[3]IRIT/INP-ENSEEIHT, Université de Toulouse, France

prénom.nom@iit.it,prénom.nom@ens-lyon.fr,prénom.nom@irit.fr

Résumé – Le problème du choix du paramètre de régularisation dans les approches variationnelles pour la résolution de problèmes inverses reste une question ouverte. Dans cette étude nous nous intéressons au problème du choix de ce paramètre, que nous noterons λ, dans la formulation du problème ℓ_2-Potts considéré pour le débruitage de signaux constants par morceaux. Le lien avec un modèle hiérarchique bayésien est mis en avant et permet de proposer une pénalisation pour le paramètre de régularisation. Une solution algorithmique permet ensuite d'estimer efficacement λ.

Abstract – The problem of choosing the regularization parameter involved in variational approaches for solving inverse problems remains an open question. In this study, we are interested in choosing this parameter, denoted λ, for the ℓ_2-Potts model which is of interest for piecewise constant signal denoising. Its link with a hierarchical Bayesian model is highlighted, it allows a penalization over λ to be derived and an algorithmic solution leads to an efficient estimation of λ.

1 Introduction

La résolution de nombreux problèmes inverses de la forme $\boldsymbol{y} = \mathcal{D}_\sigma(A\overline{\boldsymbol{x}})$ où A modélise une dégradation linéaire et $\mathcal{D}_\sigma$ une perturbation non-linéaire de paramètre $\sigma > 0$ passe souvent par la minimisation d'un critère de la forme

$$\widehat{\boldsymbol{x}}_\lambda \in \underset{\boldsymbol{x}\in\mathbb{R}^N}{\operatorname{Argmin}}\ \psi_1(A\boldsymbol{x}, \boldsymbol{y}) + \lambda\psi_2(\boldsymbol{x}) \tag{1}$$

où $\psi_1(A\cdot, \boldsymbol{y})\colon \mathbb{R}^N \to]-\infty, +\infty]$ modélise le terme d'attache aux données, $\psi_2\colon \mathbb{R}^N \to]-\infty, +\infty]$ le terme de régularisation, et $\lambda > 0$ le paramètre de régularisation. De nombreux algorithmes ont été proposés pour estimer $\widehat{\boldsymbol{x}}_\lambda$ à λ fixé dans le cadre convexe ou non convexe [1]. Dans ce travail, nous proposons d'estimer conjointement $\widehat{\boldsymbol{x}}$ et $\widehat{\lambda}$ par minimisation d'un critère de la forme

$$(\widehat{\boldsymbol{x}}, \widehat{\lambda}) \in \underset{\boldsymbol{x}\in\mathbb{R}^N, \lambda>0}{\operatorname{Argmin}}\ \psi_1(A\boldsymbol{x}, \boldsymbol{y}) + \lambda\psi_2(\boldsymbol{x}) + \phi(\lambda)$$

où ϕ est déduite à partir d'un modèle hiérarchique bayésien soigneusement choisi.

Débruitage constant par morceaux. Dans cette première étude nous revisitons le problème de débruitage constant par morceaux 1D qui consiste à détecter les changements de moyenne dans des signaux univariés. Il est rencontré dans de nombreux domaines (cf., e.g., [2, 3]) et soulève des défis de taille pour s'adapter à l'accroissement des données enregistrées par les applications modernes. Le modèle le plus couramment utilisé dans la littérature suppose que les observations $\boldsymbol{y} \in \mathbb{R}^N$ sont issues d'un signal constant par morceaux $\overline{\boldsymbol{x}}$ dégradé par un bruit blanc additif gaussien $\boldsymbol{b} \in \mathcal{N}(0_N, \sigma^2\mathbb{1}_N)$ avec $\sigma^2 > 0$ inconnue, i.e., $\boldsymbol{y} = \overline{\boldsymbol{x}} + \boldsymbol{b}$.

Ce problème a été abordé suivant différentes stratégies, comme les procédés de Cusum [2], l'inférence bayésienne hiérarchique [4, 5], ou par des formulations d'optimisation fonctionnelle impliquant soit la norme ℓ_1 [6, 7] ou la pseudo-norme ℓ_0 des premières différences du signal [8, 9].

La présente contribution s'inscrit dans ce dernier cas. Formellement, le problème considéré est un cas particulier de (1) qui consiste à trouver la solution du modèle ℓ_2-Potts, i.e.,

$$\widehat{\boldsymbol{x}}_\lambda = \underset{\boldsymbol{x}\in\mathbb{R}^N}{\operatorname{argmin}} \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{x}\|_2^2 + \lambda\|L\boldsymbol{x}\|_0 \tag{2}$$

où $L \in \mathbb{R}^{(N-1)\times N}$ est l'opérateur de premières différences, i.e., $(L\boldsymbol{x})_k = x_{k+1} - x_k$ et où $\|\cdot\|_0$ compte le nombre d'éléments non nuls. Lorsque λ est fixé, (2) peut être efficacement résolu par des algorithmes de programmation dynamique [10, 11]. Le choix de λ peut ensuite être effectué par des règles heuristiques [12, 13] ou des critères d'information pénalisant la complexité de $\widehat{\boldsymbol{x}}_\lambda$ comme les critères AIC, SIC, AICC et SICC [14]. Plus récemment, de nouveaux critères ont été proposés dans [15, 16, 17].

D'un autre côté, le cadre d'inférence bayésienne hiérarchique a reçu beaucoup d'attention pour le débruitage de signaux constants par morceaux [4, 5]. Leur intérêt réside principalement dans leur capacité à inclure les hyper-paramètres au sein du modèle bayésien et de les estimer conjointement aux paramètres d'intérêts. En contrepartie, les méthodes bayésiennes hiérarchiques ont généralement recours à des algorithmes de Monte Carlo par chaîne de Markov (MCMC), coûteux en temps de calcul, pour approcher les estimateurs bayésiens.

Dans ce travail nous proposons d'associer les méthodes de résolution efficaces du problème ℓ_2-Potts avec la flexibilité des approches bayésiennes en considérant le problème ℓ_2-Potts pénalisé suivant

$$(\widehat{\boldsymbol{x}}, \widehat{\sigma}^2, \widehat{\lambda}) \in \underset{\substack{\boldsymbol{x}\in\mathbb{R}^N, \sigma^2>0, \\ \lambda>0}}{\operatorname{Argmin}} \frac{1}{2\sigma^2}\|\boldsymbol{y}-\boldsymbol{x}\|_2^2 + \frac{\lambda}{\sigma^2}\|L\boldsymbol{x}\|_0 + \phi(\sigma^2, \lambda) \tag{3}$$

où la fonction ϕ, héritée d'un modèle hiérarchique bayésien, permet de pénaliser correctement les paramètres inconnus (σ^2, λ). La variance σ^2 est incluse dans la pénalisation car les choix de λ et σ^2 sont liés. En effet, un signal présentant les même ruptures mais un bruit d'amplitude différente nécessitera un choix de σ^2 différent.

Contributions et plan. Pour commencer, nous explicitons dans la Section 2 la forme explicite de la pénalisation ϕ à partir d'un modèle bayésien hiérarchique soigneusement choisi. Cette section décrit également une solution algorithmique permettant de résoudre (3). La Section 3 présente des résultats de simulations numériques. En particulier, nous présentons et nous examinons l'impact de plusieurs estimateurs de $\widehat{\lambda}$ et $\widehat{\sigma}^2$ sur les performances d'estimation.

2 Expression de ϕ

Reparamétrisation. Pour obtenir l'expression de ϕ, nous adoptons la stratégie suivie dans [18], proposant qu'une solution candidate constante par morceaux $\boldsymbol{x}$ peut être paramétrée par un vecteur d'indicatrices $\boldsymbol{r} \in \{0,1\}^N$ de discontinuités et un vecteur $\boldsymbol{\mu} \in \mathbb{R}^K$ des valeurs prises sur chacun des K segments. Le vecteur $\boldsymbol{r} = \left(r_i\right)_{1\leq i\leq N}$ est défini par

$$r_i = \begin{cases} 1, & \text{si il y a une discontinuité à la position } i, \\ 0, & \text{sinon,} \end{cases} \tag{4}$$

avec $r_N = 1$ afin que $K = \sum_{i=1}^N r_i$. Les valeurs prises sur chacun des segments de $\boldsymbol{x}$ sont contenues dans le vecteur $\boldsymbol{\mu} = (\mu_k)_{1\leq k\leq K}$ de telle façon que

$$(\forall k \in \{1,\ldots,K\})(\forall i \in \mathcal{R}_k) \quad x_i = \mu_k \tag{5}$$

où l'ensemble $\mathcal{R}_k \subset \{1,\ldots,N\}$ contient les indices temporels associés au $k$ième segment. Par conséquent, $\mathcal{R}_k \cap \mathcal{R}_{k'} = \{\emptyset\}$ pour $k \neq k'$ et $\cup_{k=1}^K \mathcal{R}_k = \{1,\cdots,N\}$.

Approche hiérarchique. Premièrement, compte tenu de la statistique gaussienne du bruit $\boldsymbol{b}$, la vraisemblance des observations $\boldsymbol{y}$ s'exprime

$$f\big(\boldsymbol{y}|\boldsymbol{r},\boldsymbol{\mu},\sigma^2\big) = \prod_{k=1}^{K_{\boldsymbol{r}}} \prod_{i\in\mathcal{R}_k} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\mu_k - y_i)^2}{2\sigma^2}\right). \tag{6}$$

Afin de calculer la distribution *a posteriori* des paramètres du modèle, il est nécessaire de spécifier les distributions *a priori* de $\boldsymbol{r}$, $\boldsymbol{\mu}$, et σ. Dans la littérature, il est souvent supposé que les variables $(r_i)_{1\leq i\leq N}$ sont indépendantes et identiquement distribuées selon une distribution de Bernoulli de paramètre p inconnu quantifiant la probabilité *a priori* de la présence d'un changement, indépendemment de la position [4, 19, 5], i.e, $f(\boldsymbol{r}|p) = \prod_{i=1}^{N-1} p^{r_i}(1-p)^{1-r_i}$.

D'un point de vue bayésien, un choix naturel pour la loi *a priori* de l'amplitude des segments $(\mu_k)_{1\leq k\leq K_{\boldsymbol{r}}}$ consiste à choisir des distributions gaussiennes indépendantes conjuguées $\mathcal{N}(\mu_0, \sigma_0^2)$, i.e.,

$$f(\boldsymbol{\mu}|\boldsymbol{r}) = \prod_{k=1}^{K_{\boldsymbol{r}}} \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left[\frac{-(\mu_k-\mu_0)^2}{2\sigma_0^2}\right]. \tag{7}$$

Suivant des choix couramment adoptés dans la littérature, une distribution non-informative de Jeffreys est choisie pour la variance σ^2 afin de modéliser l'absence de connaissance *a priori*, i.e., $f\left(\sigma^2\right) \propto \sigma^{-2}$, alors qu'une distribution conjuguée Beta $\mathcal{B}(\alpha_0, \alpha_1)$ est choisie pour le paramètre inconnu p :

$$f(p) = \frac{\Gamma(\alpha_0+\alpha_1)}{\Gamma(\alpha_0)\,\Gamma(\alpha_1)} p^{\alpha_1-1}(1-p)^{\alpha_0-1}. \tag{8}$$

En supposant que les paramètres $\boldsymbol{r}$, $\boldsymbol{\mu}$ et σ^2 sont *a priori* indépendants, nous pouvons exprimer la distribution *a posteriori* jointe

$$f(\boldsymbol{\Theta}|\boldsymbol{y}) \propto\ f\left(\boldsymbol{y}|\boldsymbol{r},\boldsymbol{\mu},\sigma^2\right) f\left(\boldsymbol{\mu}|\boldsymbol{r}\right) f(\boldsymbol{r}|p) f(p) f\left(\sigma^2\right). \tag{9}$$

L'idée pour établir l'expression de $\phi(\cdot)$ consiste à exhiber les similitudes entre les problèmes de minimisation (3) et la maximisation de (9). Il en résulte la Proposition suivante.

Proposition 2.1. *Pour σ_0^2 suffisamment grand (voir [20, Remarque 1]), le problème* (3) *avec*

$$\lambda = \sigma^2\left(\log\left(\frac{1-p}{p}\right) + \frac{1}{2}\log(2\pi\sigma_0^2)\right) \tag{10}$$

et

$$\begin{aligned}\phi(\sigma^2,\lambda) = {} & \frac{N}{2}\log(2\pi\sigma^2) + \log(\sigma^2) \\ & - \frac{\lambda}{\sigma^2}(N+\alpha_0-2) + \frac{N+\alpha_0-1}{2}\log(2\pi\sigma_0^2) \\ & + (N+\alpha_0+\alpha_1-3)\log\left(1+\exp\left(\frac{\lambda}{\sigma^2} - \frac{1}{2}\log(2\pi\sigma_0^2)\right)\right)\end{aligned} \tag{11}$$

coïncide avec la maximisation de (9).

Algorithme de minimisation. Comme souvent dans les méthodes d'estimation du paramètre de régularisation λ, il est proposé d'évaluer le critère pour différentes valeurs de λ définies sur une grille prédéfinie Λ tel que $(\forall\lambda\in\Lambda)$

$$(\widehat{\boldsymbol{x}}_\lambda, \widehat{\sigma}^2_\lambda) \in \underset{\boldsymbol{x}\in\mathbb{R}^N, \sigma^2>0}{\operatorname{Argmin}} \underbrace{\frac{1}{2\sigma^2}\|\boldsymbol{y}-\boldsymbol{x}\|_2^2 + \frac{\lambda}{\sigma^2}\|L\boldsymbol{x}\|_0 + \phi(\sigma^2,\lambda)}_{F(\boldsymbol{x},\sigma^2,\lambda)}. \tag{12}$$

On peut ensuite alterner la minimisation par rapport à $\boldsymbol{x}$ et σ^2. Nous pouvons noter que $\widehat{\boldsymbol{x}}_\lambda$ se réduit à la solution du problème ℓ_2-Potts (2) ne faisant pas intervenir σ^2. L'estimateur

$\widehat{\sigma}^2_\lambda$ est quant à lui obtenu comme le minimum de la fonction $F(\widehat{\boldsymbol{x}}_\lambda, \cdot, \lambda)$, i.e.,

$$\widehat{\sigma}_\lambda \in \underset{\sigma^2>0}{\text{Argmin}}\ F(\widehat{\boldsymbol{x}}_\lambda, \sigma^2, \lambda). \tag{13}$$

Finalement, nous retenons comme solution le triplet $\{\widehat{\boldsymbol{x}}_{\widehat{\lambda}}, \widehat{\sigma}^2_{\widehat{\lambda}}, \widehat{\lambda}\}$ tel que

$$\widehat{\lambda} = \text{Argmin}_{\lambda\in\Lambda} F(\widehat{\boldsymbol{x}}_\lambda, \widehat{\sigma}^2_\lambda, \lambda). \tag{14}$$

3 Expériences numériques

Impact de l'estimateur $\widehat{\sigma}^2_\lambda$ – Nous proposons de comparer deux estimateurs de $\widehat{\sigma}^2_\lambda$. Le premier est l'estimateur défini en (13) qui nécessite la résolution d'un problème de minimisation. Nous noterons cette estimée $\widehat{\sigma}^{2,\text{min}}_\lambda$. Le second est l'estimateur empirique de la variance défini comme suit

$$\widehat{\sigma}^{2,\text{emp}}_\lambda = \frac{\|\boldsymbol{y} - \widehat{\boldsymbol{x}}_\lambda\|^2}{N-1}. \tag{15}$$

Pour une réalisation $\boldsymbol{y}$ donnée et représentée sur la Fig. 1 (1ière ligne), la Fig. 1 (2ième ligne) compare $\widehat{\sigma}^{2,\text{min}}_\lambda$ (pointillés verts) et $\widehat{\sigma}^{2,\text{emp}}_\lambda$ (rouge) en fonction de λ. Les deux estimées sont très similaires à la différence près que la seconde à un comportement constant par morceaux du fait que les solutions $\{\widehat{\boldsymbol{x}}_\lambda\}_{\lambda\in\Lambda}$ sont identiques sur des intervalles de l'axe $\lambda > 0$ (nous reviendrons sur ce point au paragraphe suivant). Nous observons que les deux critères (Fig.1, 3ième ligne) ont le même comportement et partagent la même position du minimum $\widehat{\lambda}$. Par conséquent, le critère $\lambda \mapsto F(\widehat{\boldsymbol{x}}_\lambda, \widehat{\sigma}^{2,\text{emp}}_\lambda, \lambda)$ se comporte comme un excellent substitut à $\lambda \mapsto F(\widehat{\boldsymbol{x}}_\lambda, \widehat{\sigma}^{2,\text{min}}_\lambda, \lambda)$ qui nécessite de résoudre $|\Lambda|$ problèmes d'optimisation non-convexe (13). Des conclusions similaires peuvent être tirées quelque soit la réalisation $\boldsymbol{y}$. Dans la suite, nous choisirons donc l'estimateur $\widehat{\sigma}^{2,\text{emp}}_\lambda$.

Impact de la grille Λ. A présent, nous nous intéressons à l'impact du choix de la grille Λ sur la qualité des estimateurs. Nous considérons deux types de grilles. La première est la grille régulière, notée Λ^{reg}. La seconde est une grille irrégulière obtenue par la stratégie de λ-shooting introduite dans [11]. Elle réside sur l'identification des intervalles où la solution $\{\widehat{\boldsymbol{x}}_\lambda\}_\lambda$ est identique, et sélectionne une position par intervalle pour constituer la grille Λ^{shoot}. Le lecteur intéressé dispose d'une option « λ-shooting » dans le logiciel [1] et est invité à consulter [11] pour davantage de détails.

Nous observons Fig. 2 que la grille Λ^{shoot} (croix bleues) est davantage discrétisée dans la région des faibles valeurs de λ, correspondants aux solutions $\widehat{\boldsymbol{x}}_\lambda$ comportant de nombreuses discontinuités, alors que Λ^{reg} (rouge) possède plusieurs valeurs dans la région minimisant l'EQM. Cependant, puisque F n'est pas constant par morceaux, le minimum de F sur la grille Λ^{shoot} peut mener à une solution qui ne minimise pas

1. Le logiciel Matlab est disponible à l'adresse `http://perso.ens-lyon.fr/jordan.frecon/`

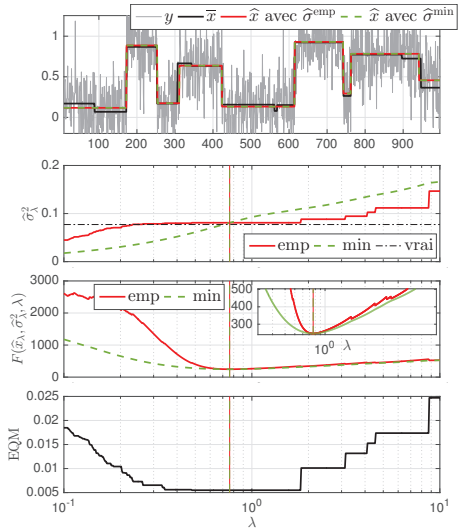


FIGURE 1 – **Comparaison entre les estimateurs $\widehat{\sigma}^2_\lambda$.** Pour l'exemple examiné $\overline{p} = 0.015$ et $\overline{\sigma}^2 = 0.08$. Les estimateurs de $\widehat{\sigma}^2_\lambda$ et les critères F associés sont respectivement comparés sur les 2ième et 3ieme lignes en fonction de λ. L'EQM entre $\widehat{\boldsymbol{x}}_\lambda$ et $\overline{\boldsymbol{x}}$ est affichée sur la 4ième ligne. L'estimée de $\widehat{\lambda}$ est indiquée sur ces figures par des traits verticaux et la solution résultante est reportée sur la 1ière ligne.

F globalement. Sur l'exemple fourni, on remarque ainsi que la solution obtenue avec Λ^{reg} minimise bien le critère F et coïncide avec la valeur minimisant l'EQM (pointillés rouges) mais que cela n'est pas le cas avec Λ^{shoot}.

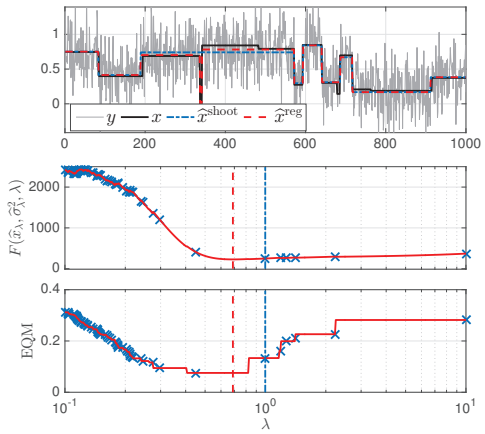


FIGURE 2 – **Illustration Λ^{reg} vs. Λ^{shoot}.** Pour l'exemple examiné $\overline{p} = 0.015$ et $\overline{\sigma}^2 = 0.08$. Le critère F et l'EQM sont respectivement comparés sur les 2ième et 3ieme lignes pour Λ^{reg} (rouge) et Λ^{shoot} (croix bleues) en fonction de λ. L'estimée $\widehat{\lambda}$ y est indiquée par des traits verticaux et la solution correspondante est affichée sur la 1ière ligne..

Nous proposons d'étudier plus en détail l'impact de la discrétisation de Λ en examinant plusieurs grilles régulières Λ^{reg} constituées respectivement de 10, 50, 100, 500 et 1000 valeurs. Les résultats de la comparaison sont présentés Fig. 3 pour deux

valeurs de $\overline{\sigma}^2$ et différentes valeurs de probabilité de rupture $\overline{p}$. Les résultats, moyennés sur 50 réalisations, montrent que les deux approches ont des performances proches avec un léger avantage pour une grille régulière avec une discrétisation suffisamment fine. Pour finir, notons que l'utilisation de Λ^{reg} mène systématiquement à un temps de calcul plus faible que Λ^{shoot} dès que $|\Lambda^{\text{reg}}| \leq 200$.

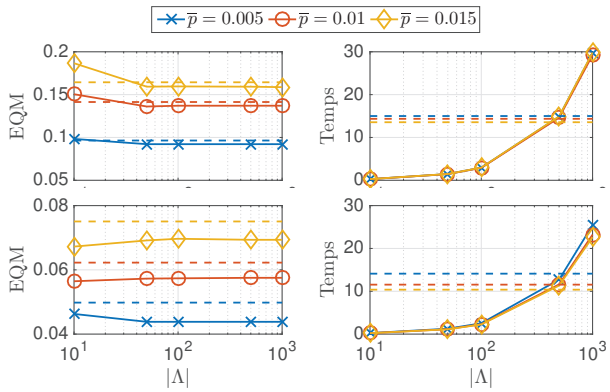


FIGURE 3 – **Impact de Λ.** Deux configurations sont examinées : $\overline{\sigma}^2 = 0.1$ (1ière ligne) et $\overline{\sigma}^2 = 0.03$ (2ième ligne) pour différentes valeurs de $\overline{p}$. Les performances, en termes d'EQM et de temps de calcul (en secondes), sont comparées pour $\Lambda = \Lambda^{\text{shoot}}$ (pointillés) et $\Lambda = \Lambda^{\text{reg}}$ (traits pleins) pour différents $|\Lambda^{\text{reg}}|$.

Comparaisons. En considérant le cas d'une grille régulière et l'estimateur empirique de $\widehat{\sigma}^{2,\text{emp}}_{\lambda}$, le bon comportement du choix de λ par cette approche est illustré Figure 4. Nous montrons en particulier l'avantage en temps de calcul par rapport à des approches de type MCMC pour des performances d'estimation similaires.

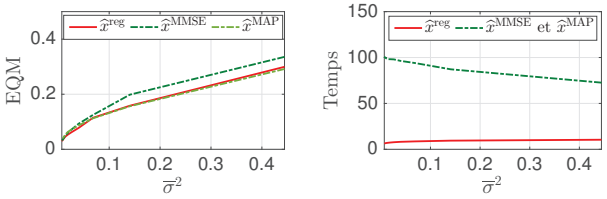


FIGURE 4 – **Comparaison avec l'état de l'art.** Dans cet exemple $\overline{p} = 0.015$ et $\overline{\sigma}^2$ varie. La méthode proposée (rouge) avec $|\Lambda^{\text{reg}}| = 500$ est comparée aux estimateurs MAP et MMSE bayésiens (vert) obtenus après 10^3 itérations.

4 Conclusions

Dans cet article, nous nous sommes intéressés à la sélection du paramètre de régularisation dans le modèle l2-Potts. On pourra également se référer à la version longue de ce travail [20]. Dans la présente contribution une attention particulière est portée sur les aspects liés à l'implémentation de la méthode. En effet, nous avons montré que le choix de la procédure algorithmique consistant à choisir une grille régulière pour Λ et l'estimateur empirique de $\widehat{\sigma}^{2,\text{emp}}_{\lambda}$ conduisent à des solutions très satisfaisantes.

Références

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

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

[3] M. Little and N. 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.

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

[5] 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, pp. 414–423, Feb. 2007.

[6] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys.D*, vol. 60, pp. 259–268, Nov. 1992.

[7] M. A. Little and N. S. Jones, "Sparse bayesian step-filtering for high-throughput analysis of molecular machine dynamics," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, (Dallas, TX), pp. 4162–4165, March 2010.

[8] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the bayesian restoration of images," *IEEE Trans. Pattern Anal. Match. Int.*, vol. 6, pp. 721–741, Jun. 1984.

[9] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems.," *Comm. Pure Applied Math.*, vol. 42, pp. 577–685, 1989.

[10] G. Winkler and V. Liebscher, "Smoothers for discontinuous signals," *Journal of Nonparametric Statistics*, vol. 14, no. 1-2, pp. 203–222, 2002.

[11] F. Friedrich, A. Kempe, V. Liebscher, and G. Winkler, "Complexity penalized M-estimation : fast computation," *J. Comput. Graph. Statist.*, vol. 17, no. 1, pp. 201–224, 2008.

[12] L. Dümbgen and K. A., "Extensions of smoothing via taut strings," *Electronic Journal of Statistics*, vol. 3, pp. 41–75, 2009.

[13] I. W. Selesnick, A. Parekh, and I. Bayram, "Convex 1-D total variation denoising with non-convex regularization,"," *IEEE Signal Process. Lett.*, vol. 22, pp. 141–144, Feb. 2015.

[14] G. Winkler, O. Wittich, V. Liebscher, and A. Kempe, "Don't shed tears over breaks," *Jahresber. Deutsch. Math.-Verein*, vol. 107, pp. 57–87, 2005.

[15] N. Zhang and D. Siegmund, "A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data," *Biometrics*, vol. 63, pp. 22–32, Mar. 2007.

[16] K. Frick, A. Munk, and H. Sieling, "Multiscale change point inference.," *Journal of the Royal Statistical Society :Series B*, vol. 76, no. 3, pp. 495–580, 2014.

[17] Z. Xia and P. Qiu, "Jump information criterion for statistical inference in estimating discontinuous curves," *Biometrika*, vol. 102, no. 2, pp. 397–408, 2015.

[18] N. Dobigeon, J.-Y. Tourneret, and M. Davy, "Joint segmentation of piecewise constant autoregressive processes by using a hierarchical model and a Bayesian sampling approach," *IEEE Trans. Signal Process.*, vol. 55, pp. 1251–1263, Apr. 2007.

[19] 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, pp. 747–758, Mar. 2002.

[20] J. Frecon, N. Pustelnik, N. Dobigeon, H. Wendt, and P. Abry, "Bayesian selection for the ℓ_2-Potts model regularization parameter : 1D piecewise constant signal denoising." arXiv preprint arXiv :1608.07739, 2016.