Consistent Change Point Detection for Piecewise Constant Signals With Normalized Fused LASSO

Arash Owrang, Mohammadreza Malek-Mohammadi, Alexandre Proutiere and Magnus Jansson

ACCESS Linnaeus Centre, School of Electrical Engineering and Computer Science
KTH Royal Institute of Technology, Sweden

Contribution

- We consider the problem of offline change point detection from noisy piecewise constant signals.
- We propose normalized fused LASSO (NFL), a modification of the fused LASSO (FL).
- We prove that NFL is consistent in detecting change points as the noise variance tends to zero.

Introduction

Piecewise constant signals play a major role in numerous applications such as comparative genomic hybridization, analysis of financial time-series, bio-medical imaging, smart power grids and data segmentation. These applications call for detecting the change points from perturbed measurements. Numerous change point detection methods have been proposed. In particular, the fused LASSO (FL) has attracted a lot of attention recently [2].

Model

We have N noisy measurements $\{y(t)\}_{t=1}^{N}$ of the true piecewise constant signal $m^*(t)$:

$$y(t) = m^*(t) + \sigma\epsilon(t) \qquad (1)$$

where $\epsilon(t) \sim \mathcal{N}(0,1)$ is temporally independent and σ is the unknown standard deviation.
The signal $m^*(t)$ is piecewise constant; i.e., there are K unknown change points $\mathcal{S} = \{s_i\}_{i=1}^{K}$ and values $\overline{m}_1, \ldots, \overline{m}_{K+1}$ such that

$$m^*(t) = \overline{m}_i; \qquad s_{i-1} < t \leq s_i$$

with the convention that $s_0 = 0$ and $s_{K+1} = N$.

Fused LASSO

FL exploits the sparsity of the derivative of the piecewise constant signal by solving [2]

$$\hat{\mathbf{m}}^{\mathrm{FL}} = \underset{\mathbf{m}\in\mathbb{R}^N}{\operatorname{argmin}} \left\{ \frac{1}{2}\|\mathbf{y}-\mathbf{m}\|_2^2 + \lambda\|\mathbf{D}\mathbf{m}\|_1 \right\}, \qquad (2)$$

where $\mathbf{y} = [y(1), \ldots, y(N)]^T$,

$$\mathbf{D} = \begin{bmatrix} -1 & 1 & 0 & \ldots & 0 \\ 0 & -1 & 1 & 0 \ldots & 0 \\ & & & \ldots & \\ 0 & \ldots & & 0 -1 & 1 \end{bmatrix}$$

is the difference operator and $\lambda > 0$ is the regularization parameter, which implicitly controls the number of detected change points.
Unfortunately, often FL fails at detecting the true change points as $\hat{\mathbf{m}}^{\mathrm{FL}}$ gets cluttered with small undesired steps forming a staircase [3].

The Failure of Fused LASSO

To understand the failure of FL, we rewrite it in the standard LASSO format

$$\min_{\mathbf{x}\in\mathbb{R}^{N-1}} \left\{ \frac{1}{2}\|\tilde{\mathbf{y}} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1 \right\}, \qquad (3)$$

where $\tilde{\mathbf{y}} = \mathbf{y} - \frac{\mathbf{y}^T\mathbf{1}}{N}\mathbf{1}$, $\mathbf{1} = [1, \ldots, 1]^T$ and $\mathbf{A} \in \mathbb{R}^{N\times(N-1)}$ is given by

$$a_{i,j} = \frac{j}{N} - 1, \text{ for } i \leq j \text{ and } \quad a_{i,j} = \frac{j}{N}, \text{ otherwise.}$$

We know that LASSO can potentially select the true non-zero elements if the irrepresentable condition holds [4]:

The Failure of Fused LASSO cont.

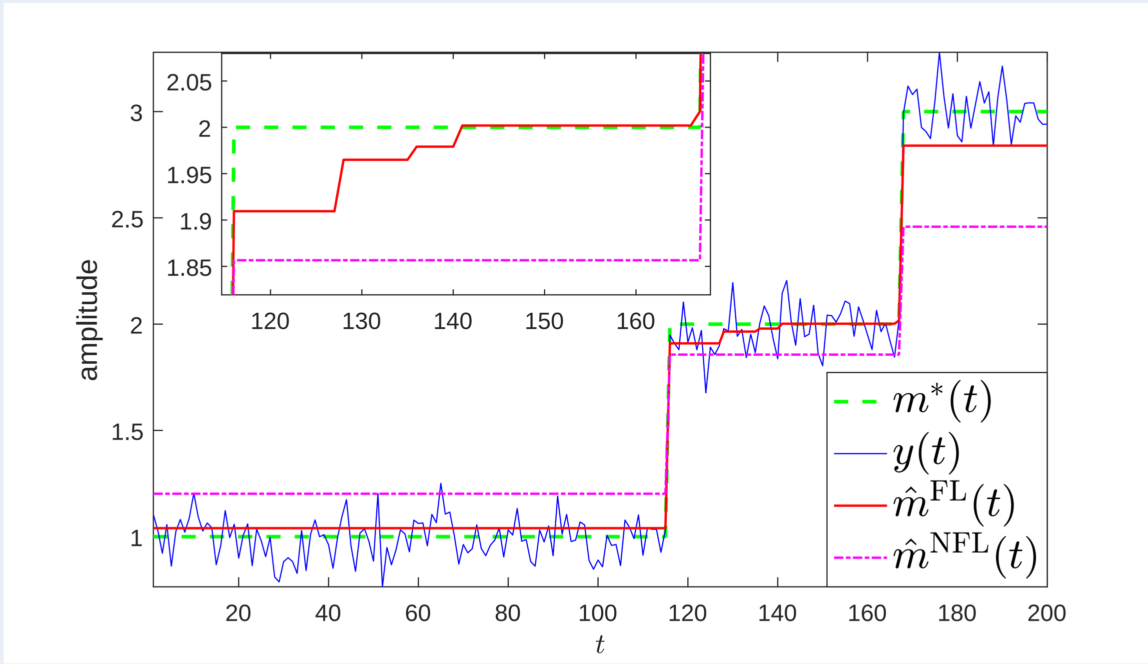

Figure: The solution of FL, $\hat{m}^{\mathrm{FL}}(t)$, is cluttered with small steps when $\sigma = 0.1$. The small box in the left top corner magnifies the intermediate level of $\hat{m}^{\mathrm{FL}}(t)$ and $\hat{m}^{\mathrm{NFL}}(t)$.

$$\nu = \|\mathbf{A}_{\mathcal{S}^c}^T\mathbf{A}_{\mathcal{S}}\left(\mathbf{A}_{\mathcal{S}}^T\mathbf{A}_{\mathcal{S}}\right)^{-1}\operatorname{sgn}(\mathbf{x}_{\mathcal{S}}^*)\|_\infty < 1, \qquad (4)$$

From the structure of the problem one can show that the irrepresentable condition is not satisfied when two consecutive steps are in the same direction [3]. On the contrary, if every two consecutive steps are in the opposite direction, i.e., $\operatorname{sgn}(x_{r-1}^*) = -\operatorname{sgn}(x_r^*)$ for all r's, the irrepresentable condition holds.

The Normalized Fused LASSO

$$\hat{\mathbf{x}}^{\mathrm{NFL}} = \underset{\tilde{\mathbf{x}}\in\mathbb{R}^{N-1}}{\operatorname{argmin}} \left\{ \frac{1}{2}\|\tilde{\mathbf{y}} - \tilde{\mathbf{A}}\tilde{\mathbf{x}}\|_2^2 + \lambda\|\tilde{\mathbf{x}}\|_1 \right\}, \qquad (5)$$

where the columns of $\mathbf{A}$ are normalized as $\tilde{\mathbf{A}} = \mathbf{A}\mathbf{W}^{-1}$ and $\mathbf{W}$ is a diagonal matrix with diagonal elements $w_{ii} = \sqrt{i(N-i)/N}$.Note that after normalizing the columns of $\mathbf{A}$, the model becomes $\tilde{\mathbf{y}} = \tilde{\mathbf{A}}\tilde{\mathbf{x}}^* + \sigma\tilde{\boldsymbol{\epsilon}}$ where $\tilde{\mathbf{x}}^* = \mathbf{W}\mathbf{x}^*$.
Alternatively, the optimization problem (5) can be transformed back to the standard FL format as

$$\min_{\mathbf{m}\in\mathbb{R}^N} \frac{1}{2}\|\mathbf{y}-\mathbf{m}\|_2^2 + \lambda\|\mathbf{W}\mathbf{D}\mathbf{m}\|_1.$$

To solve the above optimization problem, one can use the weighted taut-string algorithm , which has a computational complexity similar to that of the taut-string algorithm for solving (2).

Analysis

Lemma

Consider any valid $\mathcal{S}$.For any arbitrary $\operatorname{sgn}(\tilde{\mathbf{x}}_{\mathcal{S}}^*)$, $\tilde{\mathbf{A}}_{\mathcal{S}}$ and $\operatorname{sgn}(\tilde{\mathbf{x}}_{\mathcal{S}}^*)$ satisfy the irrepresentable condition.

Theorem

Assume that for a particular realization of $\tilde{\boldsymbol{\epsilon}}$ there is a $\lambda_p > 0$ such that

$$\|\sigma\tilde{\mathbf{A}}_{\mathcal{S}^c}^T\boldsymbol{\Pi}_{\mathcal{S}}^{\perp}\tilde{\boldsymbol{\epsilon}} + \lambda_p\tilde{\mathbf{A}}_{\mathcal{S}^c}^T\tilde{\mathbf{A}}_{\mathcal{S}}^{\dagger T}\operatorname{sgn}(\tilde{\mathbf{x}}_{\mathcal{S}}^*)\|_\infty < \lambda_p, \qquad (6)$$

$$\min_{i\in\mathcal{S}}|\tilde{x}_i^*| > \|\sigma\tilde{\mathbf{A}}_{\mathcal{S}}^{\dagger}\tilde{\boldsymbol{\epsilon}} - \lambda_p\left(\tilde{\mathbf{A}}_{\mathcal{S}}^T\tilde{\mathbf{A}}_{\mathcal{S}}\right)^{-1}\operatorname{sgn}(\tilde{\mathbf{x}}_{\mathcal{S}}^*)\|_\infty, \qquad (7)$$

where $\tilde{\mathbf{A}}_{\mathcal{S}}^{\dagger} = (\tilde{\mathbf{A}}_{\mathcal{S}}^T\tilde{\mathbf{A}}_{\mathcal{S}})^{-1}\tilde{\mathbf{A}}_{\mathcal{S}}^T$ and $\boldsymbol{\Pi}_{\mathcal{S}}^{\perp} = \mathbf{I} - \tilde{\mathbf{A}}_{\mathcal{S}}\tilde{\mathbf{A}}_{\mathcal{S}}^{\dagger}$. Then, $\hat{\mathbf{x}}^{\mathrm{NFL}}$, obtained by solving (5) with $\lambda = \lambda_p$, satisfies $\operatorname{supp}(\hat{\mathbf{x}}^{\mathrm{NFL}}) = \mathcal{S}$ and $\operatorname{sgn}(\hat{\mathbf{x}}_{\mathcal{S}}^{\mathrm{NFL}}) = \operatorname{sgn}(\tilde{\mathbf{x}}_{\mathcal{S}}^*)$.

Analysis cont.

Note that when $\tilde{\boldsymbol{\epsilon}}$ is a random vector, (6) holds with high probability for a proper choice of λ_p.
Therefore, if $\min_{i\in\mathcal{S}}|\tilde{x}_i^*|$ is large enough, NFL most likely detects the true change points.
Furthermore, as $\sigma \to 0$, (6) simplifies to the irrepresentable condition, which was shown to hold in the Lemma.
Therefore, if $\min_{i\in\mathcal{S}}|\tilde{x}_i^*|\,\Lambda_{\min}(\tilde{\mathbf{A}}_{\mathcal{S}}^T\tilde{\mathbf{A}}_{\mathcal{S}}) > \lambda_p\sqrt{K}$, NFL is asymptotically consistent in detecting $\mathcal{S}$; i.e., $\Pr(\operatorname{supp}(\hat{\mathbf{x}}^{\mathrm{NFL}}) = \mathcal{S}) \to 1$ as $\sigma \to 0$.

Empirical Results

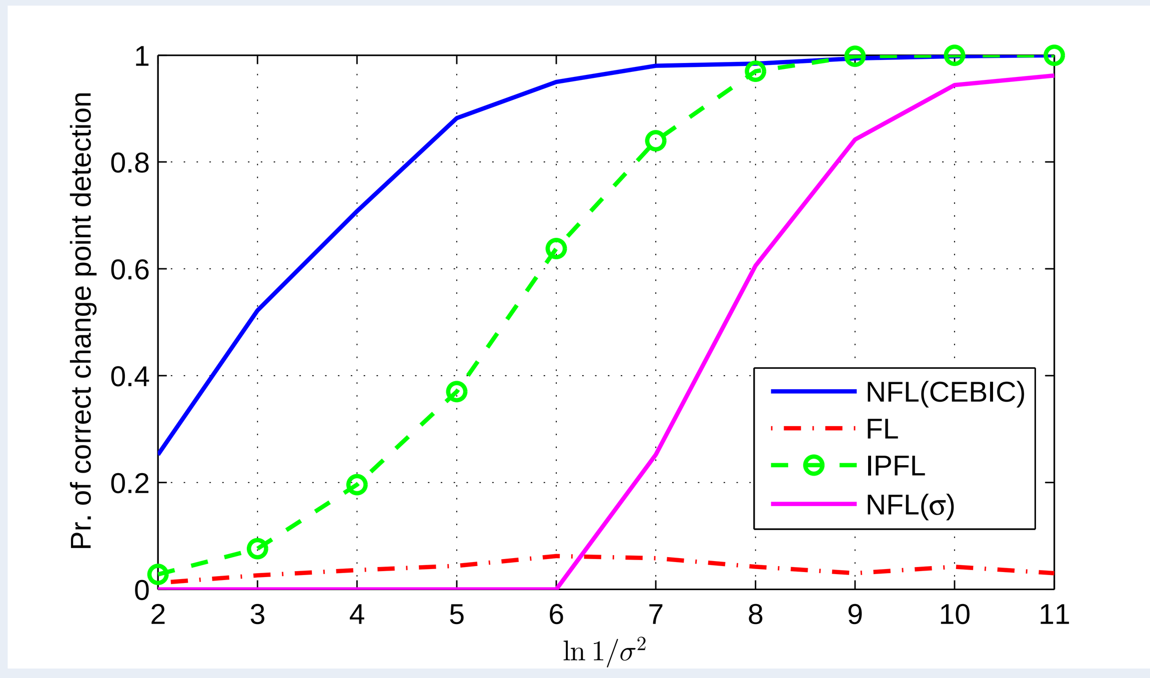

Figure: The probability of correct change point detection versus $\ln(1/\sigma^2)$ when $N = 250$ and $[\overline{m}_1, \overline{m}_2, \overline{m}_3] = [1, 2, 3]$.

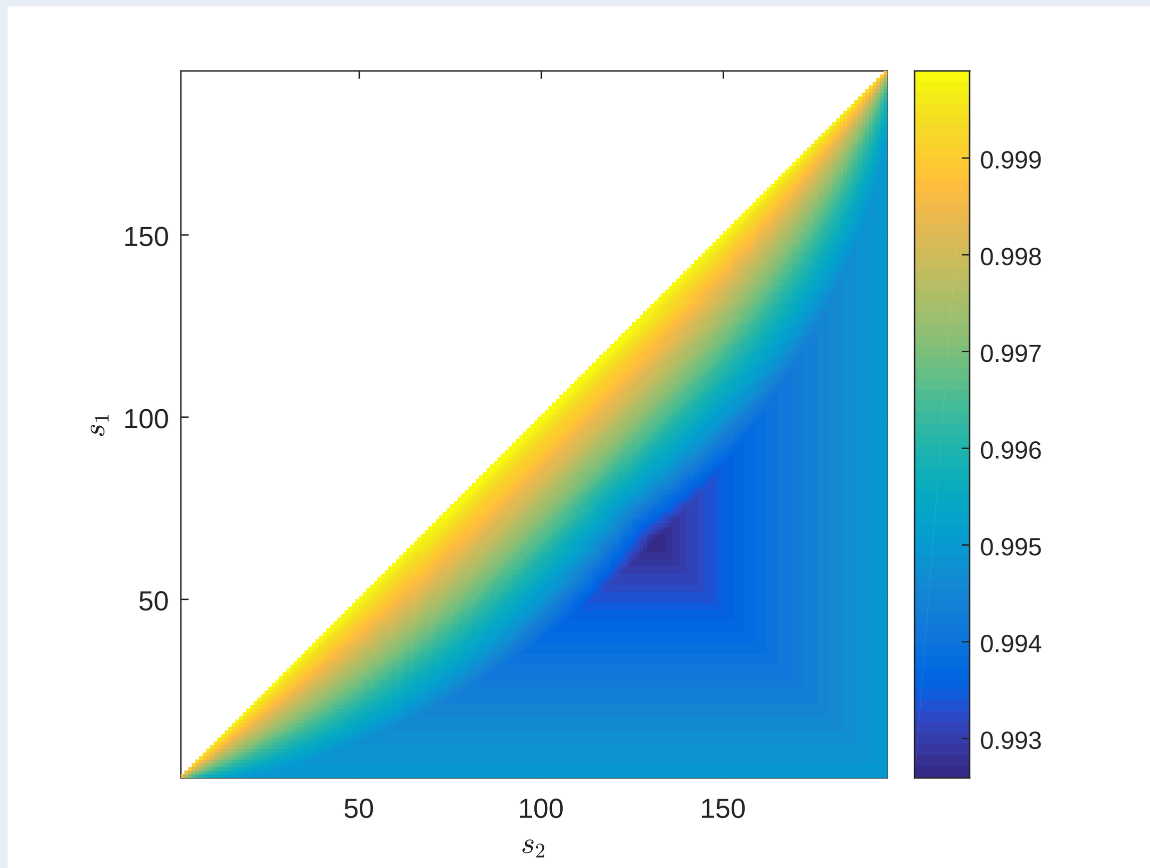

Figure: Graphical illustration of ν as a function of $\mathcal{S}$ when $K = 2$ and $N = 200$.

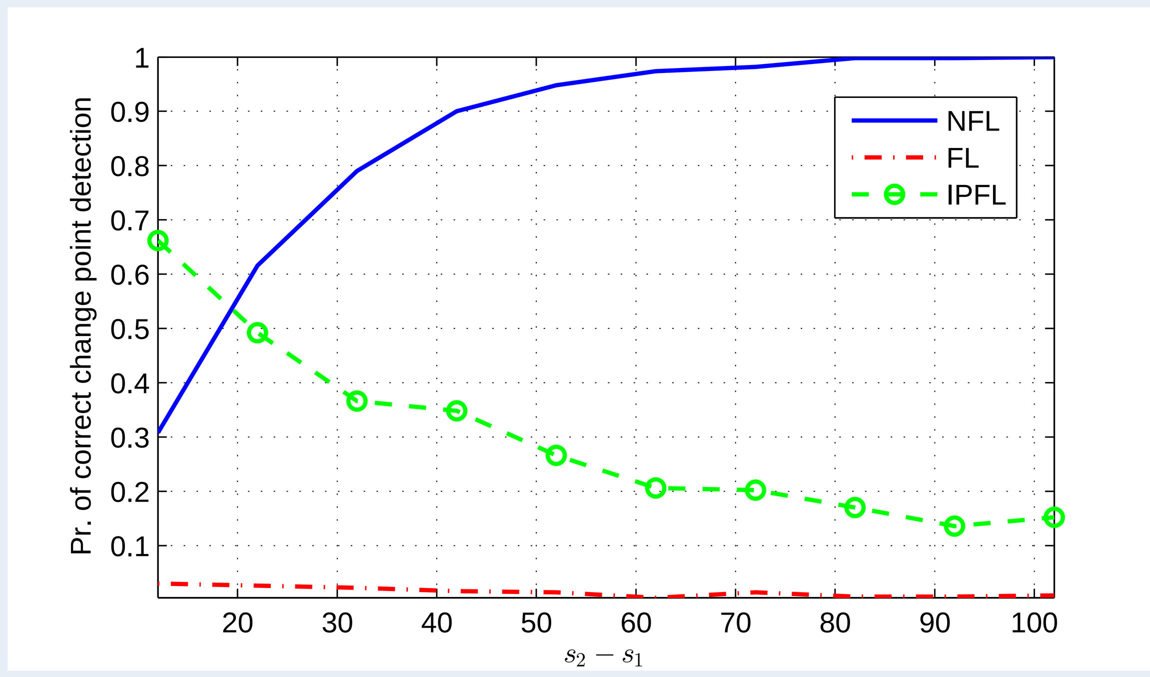

Figure: The probability of correct change point detection versus $s_2 - s_1$ when $\sigma = 0.1$, $N = 200$ and $[\overline{m}_1, \overline{m}_2, \overline{m}_3] = [1, 2, 3]$.

References

1. A. Owrang, M. Malek-Mohammadi, A. Proutiere and M. Jansson, "Consistent Change Point Detection for Piecewise Constant Signals With Normalized Fused LASSO," in IEEE Signal Processing Letters, vol. 24, no. 6, pp. 799-803, June 2017. doi: 10.1109/LSP.2017.2678516
2. R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, "Sparsity and smoothness via the fused lasso," *J. Roy. Statist. Soc.: Series B (Statistical Methodology)*, vol. 67, no. 1, pp. 91–108, 2005.
3. C. R. Rojas and B. Wahlberg, "On change point detection using the fused Lasso method," *ArXiv:1401.5408*, Jan. 2014.
4. P. Zhao and B. Yu, "On model selection consistency of lasso," *J. Mach. Learn. Res.*, vol. 7, pp. 2541–2563, Nov. 2006.