Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

On stepwise pattern recovery of the fused lasso

Q1 Junyang Qian, Jinzhu Jia *

School of Mathematical Sciences, Peking University, China
Center for Statistical Science, LMAM, LMEQF, Peking University, China

HIGHLIGHTS

- We provided necessary and sufficient conditions such that fused lasso consistently recovers the piecewise constant pattern.
- We found that in general the fused lasso is not consistent.
- We proposed a preconditioned fused lasso to overcome the non-consistent issue.
- Simulation studies support our findings.

ARTICLE INFO

Article history:
Received 8 October 2014
Received in revised form 29 June 2015
Accepted 20 August 2015
Available online xxxx

Keywords:
Fused lasso
Non-asymptotic
Pattern recovery
Preconditioning

ABSTRACT

We study the property of the Fused Lasso Signal Approximator (FLSA) for estimating a blocky signal sequence with additive noise. We transform the FLSA to an ordinary Lasso problem, and find that in general the resulting design matrix does not satisfy the irrepresentable condition that is known as an almost necessary and sufficient condition for exact pattern recovery. We give necessary and sufficient conditions on the expected signal pattern such that the irrepresentable condition holds in the transformed Lasso problem. However, these conditions turn out to be very restrictive. We apply the newly developed preconditioning method — Puffer Transformation (Jia and Rohe, 2015) to the transformed Lasso and call the new procedure the *preconditioned fused Lasso*. We give non-asymptotic results for this method, showing that as long as the signal-to-noise ratio is not too small, our preconditioned fused Lasso estimator always recovers the correct pattern with high probability. Theoretical results give insight into what controls the ability of recovering the pattern — it is the noise level instead of the length of the signal sequence. Simulations further confirm our theorems and visualize the significant improvement of the preconditioned fused Lasso estimator over the vanilla FLSA in exact pattern recovery.

1. Introduction

Assume we have a sequence of signals $(y_1, y_2, \ldots, y_n)$ and it follows the additive model Q2 Q3

$$y_i = \mu_i^* + \epsilon_i, \quad i = 1, 2, \ldots, n, \tag{1}$$

where $Y = (y_1, \ldots, y_n)^T \in \mathbb{R}^n$ is the observed signal vector, $\mu^* = (\mu_1, \ldots, \mu_n)^T \in \mathbb{R}^n$ the expected signal vector, and Q4 $\epsilon = (\epsilon_1, \ldots, \epsilon_n)^T$ the white noise such that $\epsilon_1, \ldots, \epsilon_n$ are assumed to be i.i.d. Gaussian random variables with mean 0 and variance σ^2. The model is assumed to be blocky in the sense that the signals come in blocks and have only a few change-points. To be exact, there exists a partition of $\{1, 2, \ldots, n\} = \cup_{j=1}^J \{L_j, L_j + 1, \ldots, U_j\}$ with $L_1 = 1, U_J = n, U_j \geq L_j, L_{j+1}$

* Corresponding author.
E-mail address: jzjia@math.pku.edu.cn (J. Jia).

http://dx.doi.org/10.1016/j.csda.2015.08.013
0167-9473/© 2015 Published by Elsevier B.V.

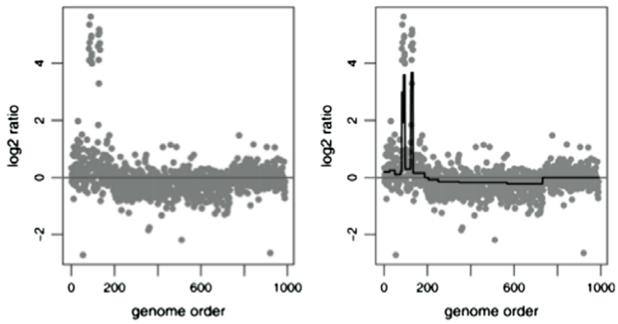

Fig. 1. This figure is from Tibshirani and Wang (2008). The fused Lasso is applied to some CGH data. The data are shown in the left panel, and the solid line in the right panel represents the estimated signals by the fused Lasso. The horizontal line is for $y = 0$.

$= U_j + 1$, and the following stepwise function holds:

$$\mu_i^* = \nu_j^* \mathbf{1}_{L_j \le i \le U_j},$$

with ν_j^*, L_j, U_j fixed but unknown. We also assume that the vector $\nu = (\nu_1, \nu_2, \ldots, \nu_J)$ is sparse, meaning that only a few of ν_j's are nonzeros. We point out that the Gaussian noise is not necessary, but we still use it to get insight of the fused lasso. The variance σ^2 of ϵ_i is the measure of noise level and does not have to be a constant here. In many cases, each observation of y_i can be an average of multiple measurements and so σ^2 decreases when the number of measurements increases. Rinaldo (2009) considers the model when $\sigma^2 = \sigma_0^2/n$, where σ_0 is a constant. We do not make this specific assumption in the development of our theory.

Featured by blockiness and sparseness, this model has many applications. For example, in tumor studies, based on the Comparative Genomic Hybridization (CGH) data, it can be used to automatically detect the gains and losses in DNA copies by taking the "signal" above as the log-ratio between the number of DNA copies in tumor cells and that in reference cells (Tibshirani and Wang, 2008). For more applications, see Tibshirani and Taylor (2011), Friedman et al. (2007) and Hoefling (2010).

One way to estimate the unknown parameters is via the Fused Lasso Signal Approximator (FLSA) defined as follows (Tibshirani et al., 2004; Friedman et al., 2007):

$$\hat{\mu}(\lambda_1, \lambda_2) = \underset{\mu}{\operatorname{argmin}} \frac{1}{2}\|Y - \mu\|_2^2 + \lambda_1\|\mu\|_1 + \lambda_2\|\mu\|_{TV}, \tag{2}$$

where $\|\mu\|_1 = \sum_{i=1}^n |\mu_i|$, $\|\mu\|_2^2 = \sum_{i=1}^n \mu_i^2$ and $\|\mu\|_{TV} = \sum_{i=1}^{n-1} |\mu_{i+1} - \mu_i|$. The L_1-norm regularization controls the sparsity (number of zeros) and the total variation seminorm ($\|\mu\|_{TV}$) regularization controls the blockiness (number of blocks or partitions).

Fig. 1 gives some CGH data, a typical example of signals with such features and a proper FLSA estimate on the data. More details and examples can be seen in Tibshirani and Wang (2008).

One important question for the FLSA is how good the estimator defined in Eq. (2) is. We analyze in this paper if the FLSA can recover the "stepwise pattern" or not. We also try to answer the following question: what do we do if the FLSA does not recover the "stepwise pattern"? To measure how good an estimator is, we introduce the following definition of Pattern Recovery.

Definition 1 (*Pattern Recovery*). An FLSA solution $\hat{\mu}(\lambda_{1n}, \lambda_{2n})$ recovers the signal pattern if and only if there exist λ_{1n} and λ_{2n}, such that

$$sign(\hat{\mu}_{i+1}(\lambda_{1n}, \lambda_{2n}) - \hat{\mu}_i(\lambda_{1n}, \lambda_{2n})) = sign(\mu_{i+1}^* - \mu_i^*), \quad i = 1, \ldots, n-1. \tag{3}$$

We use $\hat{\mu} =_{js} \mu^*$ to shortly denote (3) (*js* is the acronym for *jump sign*). The FLSA with the property of pattern recovery means that it can be used to identify both the groups and the jump directions (up or down) between groups.

The concept of pattern recovery of the FLSA is very similar to the sign recovery of the Lasso (Zhao and Yu, 2006). In fact, we will see in Section 2 that the pattern recovery property of the FLSA is equivalent to the sign recovery property of the Lasso after transformation.

For observation pairs (x_i, y_i), $i = 1, 2, \ldots, n$ with $x_i \in \mathbb{R}^p$ and $y_i \in \mathbb{R}$, the Lasso estimator is defined as follows (Tibshirani, 1996):

$$\hat{\beta}(\lambda) = \underset{\beta}{\operatorname{argmin}} \frac{1}{2}\sum_{i=1}^n (y_i - x_i^T\beta)^2 + \lambda\|\beta\|_1,$$

Please cite this article in press as: Qian, J., Jia, J., On stepwise pattern recovery of the fused lasso. Computational Statistics and Data Analysis (2015), http://dx.doi.org/10.1016/j.csda.2015.08.013