Corrections to
Properties and Refinements of the Fused Lasso

Alessandro Rinaldo

Both the statement and the proof of Theorem 2.3 in Rinaldo (2009) about the recovery properties of the fused lasso are incorrect. In this note, I will revise that statement and provide a more transparent means of analyzing the fused lasso estimator. In doing so, I will acknowledge other contributions in the literature that instead have correct claims: see, in particular, Harchaoui and Lévy-Leduc (2010), Qian and Jia (2012) and Rojas and Wahlberg (2014).

1 Introduction

We consider the data $Y = (Y_1, \ldots, Y_n)$ of the form

$$Y_i = \mu_i^0 + \epsilon_i, \quad i = 1, \ldots, n,$$

where $\epsilon_1, \ldots, \epsilon_n \overset{i.i.d.}{\sim} N(0, \sigma_n^2)$ and the unknown mean signal $\mu^0 = \mu^0(n) = (\mu_1^0, \ldots, \mu_n^0) \in \mathbb{R}^n$ is piecewise constant, i.e. there exists a set $\mathcal{J}_0 = \{i_1, i_2, \ldots, i_J\} \subset \{1, \ldots, n\}$ of J indexes with $1 < i_1 < i_2 < \ldots < i_J \leq n$ such that $\mu_i \neq \mu_{i-1}$ if and only if $i \in \mathcal{J}^0$. The coordinates $\mathcal{J}^0$ of the signal jumps are not known, nor is the vector $s^0 = (s_{i_1}, \ldots, s_{i_J}) \in \{-1, 1\}^J$ of the jump signs, where $s_1 = \mathrm{sgn}(\mu_1^0 - \mu_{i_1}^0)$ and $s_{i_l} = \mathrm{sgn}(\mu_{i_l}^0 - \mu_{i_l - 1}^0)$ for $l = 2, \ldots, J$. We further let

$$\delta = \min_{j \in \mathcal{J}^0} |\mu_j^0 - \mu_{j-1}^0|$$

denote the magnitude of the smallest jump.

We are interested in estimating the signal vector μ^0 and especially the set of jumps $\mathcal{J}^0$. Towards this end, we will use the fused lasso estimator, defined as

$$\widehat{\mu} = \widehat{\mu}(\lambda) = \mathrm{argmin}_{\mu \in \mathbb{R}^n} \frac{1}{2}\|Y - \mu\|^2 + \lambda \|\mu\|_{\mathrm{TV}}, \tag{1}$$

where λ is a positive tuning parameter and for a vector $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, $\|x\|$ denotes it Euclidean norm and $\|x\|_{\mathrm{TV}} = \sum_{i=2}^n |x_i - x_{i-1}|$ its total variation. For any $\lambda > 0$, the fused estimator is well-defined, as the solution to the problem (1) always exists and is unique.

Let $\widehat{\mathcal{J}} = \left\{\hat{i}_1, \hat{i}_2, \ldots, \hat{i}_{\hat{J}}\right\} \subset \{1, \ldots, n\}$ be the random coordinates of the jumps of the fused lasso estimator (1), taken in increasing order, where $\widehat{J} = |\widehat{\mathcal{J}}|$. That is $\widehat{\mu}_i \neq \widehat{\mu}_{i-1}$ if and only if $i \in \widehat{\mathcal{J}}$. Notice that $\hat{i}_1 > 1$ and $\hat{i}_{\widehat{J}} \leq n$. Accordingly we set $\widehat{s} = (s_{\hat{i}_1}, \ldots, s_{\hat{i}_{\widehat{J}}}) \in \{-1, +1\}^{\widehat{J}}$ to be the vector of the jumps signs of $\widehat{\mu}$, where

$$s_{\hat{i}_i} = \mathrm{sign}(\widehat{\mu}_{\hat{i}_1} - \widehat{\mu}_1) \quad \text{and} \quad s_{\hat{i}_l} = \mathrm{sign}(\widehat{\mu}_{\hat{i}_l} - \widehat{\mu}_{\hat{i}_{l-1}}), \quad l = 2, \ldots, \widehat{J}.$$

We would like to study the conditions under which the fused lasso estimator $\widehat{\mu}$ will recover the locations $\mathcal{J}^0$ of the true jumps of μ^0 and their signs s^0. Specifically, we seek to determine conditions guaranteeing that that the probability of the event of *perfect sign recovery*

$$\left\{\widehat{\mathcal{J}} = \mathcal{J}^0 \quad \text{and} \quad \widehat{s} = s^0, \quad \text{for some } \lambda > 0\right\} \tag{2}$$

tends to 1 as $n \to \infty$. To that effect, we let the size of the signal n to increase unbounded, and allow all the variables in the problem, μ^0, σ, $\mathcal{J}$, s and δ, to also change with n, though for convenience we do not make this dependence explicit in our notation.

2 The Primal and Dual Solutions

As show by Kim et al. (2009) and Tibshirani and Taylor (2011), to analyze the properties of the fused lasso solution it is very convenient to introduce explicitly the dual variables. This can be accomplished by rewriting (1) as

$$\min_{(\mu,z)\in\mathbb{R}^n\times\mathbb{R}^{n-1}} \frac{1}{2}\|Y-\mu\|^2+\lambda\|z\|_1, \quad \text{subject to } z = D\mu$$

where D is the $(n-1)\times n$ matrix

$$\begin{bmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix}, \tag{3}$$

and, for a vector $x \in \mathbb{R}^n$, $\|x\|_1 = \sum_{i=1}^n |x_i|$. The variables μ and z are called the primal and dual variable of the problems (1).

For any $\lambda > 0$, let $(\widehat{\mu}, \widehat{z}) \in \mathbb{R}^n \times \mathbb{R}^{n-1}$ be the pair of primal/dual solutions, which always exists and is unique, and whose value of course depends on the tuning parameter λ. For reasons that will become apparent below, we will let the range of the indices of the $n-1$ entries of the dual solution be $(2,\dots,n)$: that is, $\widehat{z} = (\widehat{z}_2,\dots,\widehat{z}_n)$.

Simple algebra shows that the relationship between the primal solution $\widehat{\mu} \in \mathbb{R}^n$ and the dual solution $\widehat{z} \in \mathbb{R}^{n-1}$ is given by

$$\widehat{z}_k = \begin{cases} \widehat{\mu}_1 - y_1 & j=2 \\ \widehat{\mu}_{k-1} - y_{k-1} + \widehat{z}_{k-1} & k=3,\dots,n \end{cases} \quad \text{and} \quad \widehat{z}_n = y_n - \widehat{\mu}_n,$$

which further yields the identities

$$\widehat{z}_k = \sum_{i=1}^{k-1} (\widehat{\mu}_i - y_i), k = 2,\dots,n \quad \text{and} \quad \sum_{i=1}^{n} \widehat{\mu}_i - y_i = 0. \tag{4}$$

In addition, the dual solution $\widehat{z}$ needs to satisfy the KKT conditions, which amount to the inequality

$$\max_{k=2,\dots,n} |\widehat{z}_k| \le \lambda \tag{5}$$

and the condition that

$$\widehat{\mu}_k \neq \widehat{\mu}_{k-1} \quad \text{implies} \quad \widehat{z}_k = \lambda\,\mathrm{sgn}(\widehat{\mu}_k - \widehat{\mu}_{k-1}), \qquad k=2,\dots,n. \tag{6}$$

An explicit expression for the primal solution is

$$\widehat{\mu}_i = \begin{cases} \widehat{\mu}_1 = \overline{y}(1,\hat{i}_1-1) + \frac{\lambda}{\hat{i}_1-1} s_{\hat{i}_1} & 1 \le i < \hat{i}_1 \\ \widehat{\mu}_{\hat{i}_l} = \overline{y}(\hat{i}_l,\hat{i}_{l+1}-1) + \frac{\lambda}{\hat{i}_{l+1}-\hat{i}_l}\left(s_{\hat{i}_{l+1}} - s_{\hat{i}_l}\right) & \hat{i}_l \le i < \hat{i}_{l+1}, \;\; l=1,\dots,\hat{J}-1 \\ \widehat{\mu}_{\hat{i}_{\hat{J}}} = \overline{y}(\hat{i}_{\hat{J}},n) - \frac{\lambda}{n-\hat{i}_{\hat{J}}+1} s_{\hat{i}_{\hat{J}}} & \hat{i}_{\hat{J}} \le i \le n, \end{cases} \tag{7}$$

where for a vector $x \in \mathbb{R}^n$ and integers $1 \le i \le j \le n$, we will write $\overline{x}(i,j) = \frac{1}{j-i+1}\sum_{k=i}^{j} x_k$ for the average of x over the coordinates $(i,\dots,j)$.

3 Perfect Sign Recovery

We are now ready to study the perfect sign recovery event (2).

First, we remark that with J jumps at locations $1 < i_1 < i_2 < i_J \leq n$, μ^0 is piecewise constants over $J+1$ *blocks* of contiguous coordinates of lengths $b_1, b_2, \ldots, b_{J+1}$, where

$$b_1 = i_1 - 1, \qquad b_j = i_j - i_{j-1} \text{ for } j = 2, \ldots, J, \quad \text{and} \quad b_{J+1} = n - i_J + 1.$$

Also, we set $b_{\max} = \max_{l=1,\ldots,J+1} b_l$.

Noiseless Recovery

It is instructive to first focus on perfect sign recovery in the noiseless setting when $\epsilon = 0$, so that $y = \mu$. Though clearly uninteresting, the analysis of this case is illustrative and also leads to the same "incoherence" results established in the literature (see (see Harchaoui and Lévy-Leduc, 2010; Qian and Jia, 2012; Rojas and Wahlberg, 2014) in a rather straightforward way, as illustrated below in Section 5.

It is not hard to see that, for any $\lambda < \frac{\delta b_{\min}}{4}$, the fused lasso solution yields perfect sign recovery. Indeed, for all such λ, by (4) and (7) the following is a pair of primal and dual solutions satisfying $\widehat{z}_k = \lambda s_k$ if $k \in \mathcal{J}^0$ and $|\widehat{z}_k| \leq \lambda$ otherwise (in fact, the inequality is strict):

$$\widehat{\mu}_i = \begin{cases} \mu_1 + \frac{\lambda}{i_1 - 1} s_{i_1} & 1 \leq i < i_1 \\ \mu_{i_l} + \frac{\lambda}{i_{l+1} - i_l} \left(s_{i_{l+1}} - s_{i_1} \right) & i_l \leq i < i_{l+1}, \quad l = 1, \ldots, J-1 \\ \mu_{i_J} - \frac{\lambda}{n - i_J + 1} s_{i_J} & i_J \leq i \leq n \end{cases}$$

and

$$\widehat{z}_k = \begin{cases} \lambda s_k & k \in \mathcal{J} \\ \frac{k-1}{i_1 - 1} \lambda s_{i_1} & 2 \leq k < i_1 \\ \lambda s_{i_l} + \frac{k - i_l}{i_{l+1} - i_l} \lambda \left(s_{i_{l+1}} - s_{i_l} \right) & i_l < k < i_{l+1}, \quad l = 1, \ldots J-1 \\ \lambda s_{i_J} - \lambda \frac{k - i_J}{n - i_J + 1} s_{i_J} & i_J < k \leq n. \end{cases} \tag{8}$$

The previous expression shows that dual solution in the noiseless case is a piecewise linear function over $(2, \ldots, n)$ with knots at the coordinates in $\mathcal{J}^0$, where it takes the values $\pm\lambda$. The slope of this function is $\frac{\lambda}{i_1 - 1} s_{i_1}$ for $k = 2, \ldots, i_1 - 1$ and $-\frac{\lambda}{n - i_J + 1} s_{i_J}$ for $k = i_J + 1, \ldots, n$. For $i_l < k < i_{l+1}$, with $l = 1, \ldots, J-1$, the slope is

$$\begin{cases} \frac{2\lambda}{i_{l+1} - i_l} & \text{if} \quad s_{i_{l+1}} = 1, s_{i_l} = -1 \\ -\frac{2\lambda}{i_{l+1} - i_l} & \text{if} \quad s_{i_{l+1}} = -1, s_{i_l} = 1 \\ 0 & \text{if} \quad s_{i_{l+1}} = s_{i_l}. \end{cases}$$

In particular, if two consecutive jumps are to occur in the same direction, i.e. if they are to form what Rojas and Wahlberg (2014) refer to as a *staircase* pattern, the value of the dual variables at the coordinates in between the jumps is constant and equal to λ or $-\lambda$, depending whether the jumps are upwards or downwards, respectively. This suggests that in presence of a staircase, even a minuscule amount of noise would violate the KKT conditions along those coordinates. For non-staircase scenarios, the value of the dual variables in a neighborhood of a jump coordinate, say i_l with $2 < l \leq J$, will converge to the boundary values $\pm\lambda$ as b_l or b_{l+1} grow unbounded. This also suggest that the presence of non-vanishing noise will lead to the violation the KKT conditions in a neighborhood of the locations of the true jumps. In Section 5 below, we recast these findings in the form of an incoherence conditions for the fused lasso.

Noisy Recovery

Analogous calculations based on the identities (4) and (7) show that when, the noise variables $(\epsilon_1, \ldots, \epsilon_n)$ are present, perfect sign recovery (2) will occur for some $\lambda < \frac{\delta b_{\min}}{4}$ if and only if the dual solutions $\widehat{z} =$

$(\widehat{z}_2, \ldots, \widehat{z}_n)$ are of the form

$$\widehat{z}_k = \begin{cases} \lambda s_k & k \in \mathcal{J} \\ (k-1)\left[\frac{\lambda s_{i_1}}{i_1 - 1} + \overline{\epsilon}(1, i_1 - 1) - \overline{\epsilon}(1, k-1)\right] & 1 < k < i_1 \\ \lambda s_{i_l} + (k - i_l)\left[\frac{\lambda}{i_{l+1} - i_l}\left(s_{i_{l+1}} - s_{i_l}\right) + \overline{\epsilon}(i_l, i_{l+1} - 1) - \overline{\epsilon}(i_l, k-1)\right] & i_l < k < i_{l+1}, \quad l = 1, \ldots J-1 \\ \lambda s_{i_J} - (k - i_J)\left[\frac{\lambda}{n - i_J + 1} s_{i_J} + \overline{\epsilon}(i_J, n) - \overline{\epsilon}(i_J, k-1)\right] & i_J < k \leq n, \end{cases} \tag{9}$$

and satisfy the KKT conditions (5) and (6). In fact, using Lemma 1 in Tibshirani and Taylor (2011), we see that, almost surely with respect to the distribution of $(\epsilon_1, \ldots, \epsilon_n)$, $|\widehat{z}_k| < \lambda$ for all $k \notin \widehat{\mathcal{J}}$.

Comparing the previous expression with equation (8) describing the form of the dual solution in the noiseless setting, it becomes apparent that the presence of noise results in additional stochastic terms, one for each block, involving partial averages of the errors. As it turns out, these terms are discrete Brownian bridges. By the standard discrete Brownian bridge of length $N \in \mathbb{N}$, denoted with $\{B_N(j), j = 1, \ldots, N\}$, we mean the centered Gaussian stochastic process on $\{1, \ldots, N\}$ with covariance function given by

$$\mathbb{E}\left[B_N(j) B_N(j')\right] = \min\{j, j'\} - \frac{j\, j'}{N}, \quad 1 \leq j, j' \leq N.$$

The coordinate values of the process B_N can be expressed using partial averages of a sequence of N independent Gaussian variables. Specifically,

$$B_N(j) = \sum_{i=1}^{j} \eta_i - j\overline{\eta}(1, N) = j\left(\overline{\eta}(1, j) - \overline{\eta}(1, N)\right), \quad j = 1, \ldots, N,$$

where $\{\eta_1, \ldots, \eta_N\}$ are i.i.d. standard Gaussian variates. Thus, the joint distribution of $\{B_N(j), j = 1, \ldots, N\}$ coincides with the joint conditional distribution of the random walk $\{\sum_{i=1}^{j} \eta_j, j = 1, \ldots, N\}$ given $\sum_{i=1}^{N} \eta_i = 0$. It is also immediate to see that the discrete Brownian bridge is symmetric and possesses the time reversal property: both $-B_N$ and $\{B_N(N-j), j = 1 \ldots, N\}$ are discrete Brownian bridges whenever B_N is. See, e.g., Anderson and Stephens (1997).

Therefore, using (9), we conclude that, whenever exact sign recovery occurs for some $\lambda > 0$, then the dual solution $\widehat{z}$ will be comprised of $J+1$ consecutive discrete Brownian bridges with drifts of lengths $b_1 + 1, \ldots, b_{J+1}$ and jointing at the coordinates $\mathcal{J}^0$, where they take the values $\{\lambda s_k, k \in \mathcal{J}^0\}$. The drift terms are given by the values of the dual solution at the same λ in the noiseless case (8). Furthermore, in order to satisfy the condition (5), all these Brownian bridges must take values inside the interval $[-\lambda, \lambda]$. Using the symmetry and time reversal property of the standard Brownian bridge, we can express these conditions equivalently as the conditions that independent Brownian bridges, one for each block, will not cross certain linear boundaries.

Lemma 3.1. *The probability that fused lasso will yield perfect sign recovery* (2) *is the probability that* $J+1$ *independent standard discrete Brownian bridges*

$$B_{b_1+1}, B_{b_2}, \ldots, B_{b_J+1},$$

will satisfy simultaneously the constraints

- *first jump constraint:*

$$-\frac{\lambda}{\sigma} - j\frac{\lambda}{\sigma(i_1 - 1)} \leq B_{b_1+1}(j) \leq \frac{\lambda}{\sigma} - j\frac{\lambda}{\sigma(i_1 - 1)}, \quad j = 1, \ldots, b_1;$$

- *last jump constraint:*

$$-\frac{\lambda}{\sigma} - j\frac{\lambda}{\sigma(n - i_J + 1)} \leq B_{b_{J+1}}(j) \leq \frac{\lambda}{\sigma} - j\frac{\lambda}{\sigma(n - i_J + 1)}, \quad j = 1, \ldots, b_{J+1} - 1;$$

- *intermediate jump constraints: for $l = 1, \ldots, J-1$, if $s_{i_l} \neq i_{s_{l+1}}$*

$$-j \frac{2\lambda}{\sigma(i_{l+1} - i_l)} \leq B_{b_l}(j) \leq \frac{2\lambda}{\sigma} - j \frac{2\lambda}{\sigma(i_{l+1} - i_l)}, \quad j = 1, \ldots, b_l - 1,$$

while if $s_{i_l} = s_{i_{l+1}}$ (a staircase block)

$$B_{b_l}(j) \geq 0, \quad j = 1, \ldots, b_l - 1. \tag{10}$$

We remark that the relevance of the discrete Brownian bridge and, more generally, of the fluctuations of partial sums of random sequences for estimation and testing in change detection problem is certainly not new. See, e.g., Siegmund (1986), Yao and Au (1989), Lavielle and Moulines (2000) and references therein. With regards to the performance of the fused lasso, Rojas and Wahlberg (2014) first noted that the dual solution to the fused lasso problem take the form of a sequence of discrete Brownian bridges.

We can now rely on the above representation to identify those conditions, if any, under which the event of perfect sign recovery (2) has probability tending to one. Towards that end, we must distinguish between two different signal patterns.

The Staircase Blocks

For a staircase block, (10) shows the probability that the KKT conditions are violated is the same as the probability that a standard Brownian bridge of length b_l ever crosses the zero line. By Andersen (1953), this probability is exactly $\frac{b_l - 1}{b_l} \geq 1/2$, tending to 1 as $b_l \to \infty$. It is important to also notice that, this probability is independent of the level of the variance σ^2 and the choice of regularization parameter λ. Thus, **for staircase blocks, perfect sign recovery is impossible**.

The Non-staircase Blocks

We now turn to the case of a non-staircase block, say block l. For brevity, we set $a_l = \frac{\lambda}{\sigma_n b_l}$.

Notice first that if $a_l = O(1)$, then there is a non-vanishing probability that the KKT conditions will be violated, and, therefore, **perfect sign recovery is not possible**. To get a crude lower bound on this probability, it is enough to observe that is is at least the probability that $B_{b_l}(1) \geq 2a_l$, which is bounded away from zero under the assumed scaling of a_l. However, it is worth observing that the violation of the KKT conditions happens near the boundaries of the block. In fact, let $f(b_l)$ be an increasing function of b_l such that $f(b_l) \to \infty$ with $f(b_l) = o(b_l)$ and $f(b_l) - \log(b_l - f(b_l)) \to \infty$ as $b_l \to \infty$, for instance b_l^c, for any $0 < c < 1$. Then, using a simple Gaussian tail bound and the union bound, the probability that $B_{b_l}(j) \geq 2a_l j$ for some $j \geq f(n)$ is bounded by

$$(b_l - f(b_l)) \exp\left\{-2a_l^2 f(b_l)\right\}$$

which vanishes provided that $a_l = O1)$. Therefore, with probability tending to one as $b_l \to \infty$, the KKT conditions may only be violated for the coordinates $j < f(b_l)$.

Now let us suppose that a_l is allowed to grow unbounded. Then, using again standard Gaussian tail bounds, the union bound and the fact that $\frac{a_l^2 j^2}{j(1-j/b_l)} \geq a_l^2$ for all $j = 1, \ldots, b_l - 1$, we obtain that

$$\mathbb{P}\left(B_{b_l}(j) \geq 2a_l j, \text{ for some } j = 1, \ldots, b_l\right) \leq \sum_{j=1}^{b_l - 1} \exp\left\{-\frac{4a_l^2 j^2}{2j(1-j/b_l)}\right\} \leq b_l \exp^{-2a_l^2}.$$

By setting $a_l = \sqrt{\frac{1}{2}\log(n^c b_l)}$, for any $c > 0$, this probability is $1/n^c$. By symmetry, the probability that the KKT conditions are violated for block l is no larger than $2/n^c$. Letting J_a denote the number of non-staircase blocks, we obtain that the probability that the KKT conditions are violated for the non-staircase blocks is no larger than $2/n^c$ whenever, in addition to $\lambda > \frac{\delta b_{\min}}{4}$,

$$\frac{\lambda}{\sigma} \geq b_{\max}\sqrt{\frac{1}{2}\log(n^c J_a b_{\max})}. \tag{11}$$

Thus, **for non-staircase blocks perfect sign recovery holds under the above conditions**.

4 Discussion

In this note, I have shown that the probability that the fused lasso delivers perfect sign recovery for the coordinates jumps of a piecewise constant signal is non-vanishing for signals with staircase patterns and can be vanishing otherwise if (11) holds. This provides a correction to the wrong claims made in Rinaldo (2009).

Clearly, the condition (11) (along with the condition $\lambda < \frac{\delta b_{\min}}{4}$) is extremely strong. Rojas and Wahlberg (2014) have considered a weaker notion of approximate sign recovery for the fused lasso that allows for mistakes to be made in coordinates close to $\mathcal{J}^0$.

5 Appendix: The Incoherence Condition for the Fused Lasso

In this section we recast the impossibility and difficulties of perfect sign recovery by the fused lasso estimator in the form of an incoherence condition, as is customary in the analysis of lasso type problem. See, in particular, Harchaoui and Lévy-Leduc (2010); Qian and Jia (2012); Rojas and Wahlberg (2014).

For the sake of notational consistency, we will index the $n-1$ rows of the matrix D, defined in (3), with $\{2,\ldots,n\}$. For the subset $\mathcal{J}^0$ of $\{2,\ldots,n\}$ indexing the J coordinate jumps of the signal μ^0, we will write $D_{-\mathcal{J}^0}$ for the matrix of dimension $(n-1-J)\times n$ obtained by removing the rows of D corresponding to the indexes in $\mathcal{J}^0$. Likewise, $D_{\mathcal{J}^0}$ is the $J \times n$ submatrix of D corresponding to the rows indexed by $\mathcal{J}^0$. We use an analogous notations for the vector of the dual solutions $\widehat{z} \in \mathbb{R}^{n-1}$: we will write $\widehat{z}_{\mathcal{J}^0} \in \mathbb{R}^J$ and $\widehat{z}_{-\mathcal{J}^0} \in \mathbb{R}^{n-1-J}$ for the vectors obtained by considering only the entries of $\widehat{z}$ in coordinates $\mathcal{J}^0$ and $\{2,\ldots,n\}\setminus\mathcal{J}^0$, respectively.

If $\widehat{\mu}$ is to yield perfect sign recovery, the KKT conditions imply that the dual solution $\widehat{z}$ must satisfy

$$\widehat{z}_{\mathcal{J}^0} = \lambda \mathrm{sgn}\left(D_{\mathcal{J}^0}\mu^0\right) = \lambda s$$

and using the second equation on page 20 of Tibshirani and Taylor (2011),

$$\widehat{z}_{-\mathcal{J}^0} = \left(D_{-\mathcal{J}^0}D^\top_{-\mathcal{J}^0}\right)^{-1} D_{-\mathcal{J}^0}\epsilon - \left(D_{-\mathcal{J}^0}D^\top_{-\mathcal{J}^0}\right)^{-1} D_{-\mathcal{J}^0}\left(\lambda D^\top_{\mathcal{J}^0}s\right), \tag{12}$$

with $\|\widehat{z}\|_\infty \leq \lambda$.

A routine application of the primal-dual witness type-arguments (Wainwright, 2009) further yields that the incoherence-type of condition for the lasso problem is

$$\left\|\left(D_{-\mathcal{J}^0}D^\top_{-\mathcal{J}^0}\right)^{-1} D_{-\mathcal{J}^0}\left(D^\top_{\mathcal{J}^0}s\right)\right\|_\infty < 1-\kappa, \tag{13}$$

for a fixed $\kappa \in (0,1)$ and for all n large enough.

Now we show how the previous expression is related to the simpler formulas (8). It can be verified with simple yet tedious calculations that the matrix $\left(D_{-\mathcal{J}^0}D^\top_{-\mathcal{J}^0}\right)^{-1} D_{-\mathcal{J}^0}$ is the block diagonal matrix

$$\begin{bmatrix} D_1 & 0 & 0 & \ldots & 0 \\ 0 & D_2 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & D^{J+1} \end{bmatrix},$$

where, for $l = 1, \dots, J+1$,

$$D_l = -\left(A_{b_l} + B_{b_l} E_{b_l}\right).$$

In the above expression, A_{b_l} has dimension $(b_l - 1) \times b_l$ and is of the form

$$\begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 1 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & \dots & 1 & 0 \end{bmatrix},$$

E_{b_l} is the $(b_l - 1) \times b_l$ matrix whose entries are all 1's and B_{b_l} the $(b_l - 1)$-dimensional diagonal matrix with diagonal entries $\left(\frac{1}{b_l}, \frac{2}{b_l}, \dots, \frac{b_l - 1}{b_l}\right)$. Explicitly, for $x \in \mathbb{R}^{b_l}$, the j-th coordinate of the $(b_l - 1)$-dimensional vector $\left(A_{b_l} + B_{b_l} E_{b_l}\right) x$ is

$$\sum_{i=1}^{j} x_i + \frac{j}{b_l} \sum_{i=1}^{b_l} x_i = j\left(\overline{x}(1, j) - \overline{x}(1, b_l)\right), \quad j = 1, \dots, b_l - 1. \tag{14}$$

Similar calculations appear also in Harchaoui and Lévy-Leduc (2010).

We now take note that the n-dimensional vector $D^\top_{\mathcal{J}^0} s$ is the concatenation of $J+1$ vectors of lengths $b_1, \dots, b_{J+1}$ such that

- the first subvector of length b_1 has zero entries along each of its coordinate, except the last coordinate, which takes the value $-s_1$;
- the l-subvector of length b_l, for $l = 2, \dots, J$, has zero entries along each of its coordinate, except the first and last coordinates, which take the values s_{l-1} and $-s_l$, respectively;
- the last subvector of length b_{J+1} has zero entries along each of its coordinates, except the first coordinate, which takes the value s_J.

In light of this fact and (14), it follows that the first term on the right hand side of equation (12) matches formula (9) and the second term matches (8), for all $k \notin \mathcal{J}^0$. In particular, the incoherence condition (13) requires the values of the dual solution $\widehat{z}$ in the coordinates outside $\mathcal{J}$ to be bounded away from $\pm\lambda$. As clearly seen from (8), this cannot happen in a staircase scenario (where the values of the dual solution are $\pm\lambda$) and can only be verified in non-staircase jumps only provided that the size of the block remain bounded.

References

Andersen, S. E. (1953). On the fluctuations of sums of random variables, *Mathematica Scandinavica,* 1, 263–285.

The continuous and discrete Brownian bridges: Representations and applications, *Linear Algebra and its Applications,* 264, 145–171.

Harchaoui, Z. and Lévy-Leduc, C. (2010). Multiple Change-Point Estimation With a Total Variation Penalty *Journal of the American Statistical Association*, 105(492), 1480–1493.

Kim, S.-J., Koh, K. Boyd, S. and Gorinevsky, D. (2009). ℓ_1 Trend Filtering, SIAM Review, 51 (2), 339–360.

Lavielle, M., and Moulines, E. (2000). Least-Squares Estimation of an Unknown Number of Shifts in a Time Series, *Journal of Time Series Analysis,* 21 (1), 33–59.

Qian, J. and Jia, J. (2012). On Pattern Recovery of the Fused Lasso, available at http://arxiv.org/pdf/1211.5194v1.pdf

Rinaldo, A. (2009). Properties and Refinements of the Fused Lasso, *Annals of Statistics,* 5B, 2922–2952.

Rojas, C. R. and Wahlberg, B. (2014). On Change Point Detection Using the Fused Lasso Method, available at http://arxiv.org/pdf/1401.5408v1.pdf

Siegmund, 1986. Boundary Crossing Probabilities and Statistical Applications, *Annals of Statistics,* 14(2). 361–404.

Tibshirani, R.J. and Taylor, J. (2011). The solution path of the generalized lasso, *Annals of Statistics,* 39(3), 1335–1371.

Yao, Y., and Au, S. T. (1989). Least-Squares Estimation of a Step Function, *Sankhya,* Ser. A, 51(3), 370–381.

Wainwright, M. (2009) Sharp Thresholds for High-Dimensional and Noisy Sparsity Recovery Using ℓ_1-Constrained Quadratic Programming (Lasso), IEEE Transactions on Information Theory, 55(5), 2183–2201.