On the Complexity of the Weighted Fused Lasso

José Bento   jose.bento@bc.edu   Ralph Furmaniak   rcf@acm.org   Surjyendu Ray   raysc@bc.edu

Abstract—The solution path of the 1D fused lasso for an n-dimensional input is piecewise linear with at most $\mathcal{O}(n)$ segments [1], [2]. However, existing proofs of this bound do not hold for the *weighted* fused lasso. At the same time, results for the generalized lasso, of which the weighted fused lasso is a special case, allow more than $\Omega(3^n)$ segments [3]. In this paper, we prove that the number of segments in the solution path of the weighted fused lasso is $\mathcal{O}(n^2)$, and that, for some instances, it is $\Omega(n^2)$. We also give a new, very simple, proof of the $\mathcal{O}(n)$ bound for the fused lasso.

Index Terms—Filter, Lasso, Projection, Proximal Operator, Sum of Absolute Differences, Total Variation, Weights

I. INTRODUCTION

The generalized lasso solves

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \ \frac{1}{2}\|y - Ax\|_2^2 + \gamma\|Dx\|_1, \qquad (1)$$

where $\gamma \geq 0$, $D \in \mathbb{R}^{p \times n}$, $A \in \mathbb{R}^{m \times n}$ and $y \in \mathbb{R}^m$. A special case of this problem, which is important for signal processing (see [4] and references therein for applications), is

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \ \frac{1}{2}\sum_{t=1}^{n}(x_t - y_t)^2 + \gamma\sum_{t=1}^{n-1}\alpha_t|x_{t+1} - x_t|. \qquad (2)$$

We call (2) the *weighted 1-D fused lasso* (W1FL) with input y and weights $\alpha_t \geq 0 \ \forall t$, and we distinguish it from the well studied special case when $\alpha_t = 1 \ \forall t$, which is known as the 1-D fused lasso (1FL).

There are efficient algorithms to solve 1FL for a fixed γ. *Direct algorithms*, algorithms that terminate with an exact solution, include the Taut String algorithm, [5], and the algorithm of [6], based on dynamic programing, both with a worst case complexity of $\mathcal{O}(n)$; and the algorithm of [7], which is very fast in practice, but has $\mathcal{O}(n^2)$ worst case complexity. This algorithm has recently been improved to finish in $\mathcal{O}(n)$ steps, [8]. There are also algorithms that can deal with W1FL, for fixed γ, in $\mathcal{O}(n^2)$ iterations, [9], and in $\mathcal{O}(n)$ iterations, [10]. *Iterative algorithms*, mostly first-order fixed-point methods, include [11]–[19]. Some of these are based on the ADMM method, known to achieve the fastest possible convergence rate among all first order methods, [20], [21]. However, in many applications, when precision is crucial, or when implementing a termination procedure has a non-negligible computational cost, direct algorithms are preferred.

Frequently, we are not just interested in solving (2) for a single γ. Let $x^*(\gamma)$ be the unique solution of (2). There are degenerate cases where (1) does not have a unique solution, [22]. However, our problem has a unique solution because (2) is strictly convex. An important problem is characterizing the set $\{x^*(\gamma) : \gamma \geq 0\}$, known as the *solution path* of (2). This might be necessary, for example, to efficiently "tune" the 1FL, i.e., find the value of γ that gives the best result in a given
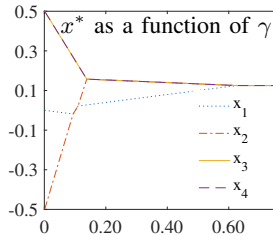
If $\alpha_1 = 1/50$, $\alpha_2 = 1/2$, $\alpha_3 = 1/2$, $y_1 = 0$, $y_2 = -1/2$, $y_3 = 1/2$ and $y_4 = 1/2$ then $x_1^* = x_2^*$ when $\gamma \in [\frac{25}{27}, \frac{25}{23}] \cup [\frac{25}{4}, \infty)$ and $x_1^* \neq x_2^*$ otherwise. At $\gamma = \frac{25}{23}$, variables x_1 and x_2 "un-fuse" and at $\gamma \in \{\frac{25}{27}, \frac{25}{4}\}$ they "fuse".

Fig. 1. Example showing that in W1FL, variables can "fuse" and "un-fuse". Hence, existing proofs that $T = \mathcal{O}(n)$ do not hold for W1FL.

application, [23], [24]. Another example is when we want to use a W1FL-path-solver to find the unique solution $\tilde{x}^*(\tilde{\gamma})$ of

$$\underset{\tilde{x} \in \mathbb{R}^n}{\min} \ \frac{1}{2}\sum_{t=1}^{n}(\tilde{x}_t - y_t)^2 \text{ subject to } \sum_{t=1}^{n-1}\alpha_t|\tilde{x}_{t+1} - \tilde{x}_t| \leq \tilde{\gamma}. \quad (3)$$

One approach is to see, cf. [25], that $\tilde{x}^*(\tilde{\gamma}) = x^*(\gamma)$ if

$$\gamma = \max_i |y_i - \tilde{x}_i^*(\tilde{\gamma})| \text{ or inversely if } \tilde{\gamma} = \|x^*(\gamma)\|_1. \quad (4)$$

We can then use the path $\{x^*(\gamma)\}$, and (4), to find which γ we should use in our solver to get $\tilde{x}^*(\tilde{\gamma})$. Relations (4) show that finding the solution paths of (3) and of (2) is equivalent.

Characterizing $\{x^*(\gamma)\}$ is possible because $x^*(\gamma)$ is a continuous piecewise linear function of γ, with a finite number T of different linear segments, a result that follows directly from the KKT conditions, [2]. Therefore, to characterize $x^*(\gamma)$, we only need to find the *critical values* $\{\gamma_i\}_{i=1}^{T-1}$ at which $x^*(\gamma)$ changes linear segment, and the value of $x^*(\gamma)$ at these γ's.

All efficient existing algorithms that find the solution path $\{x^*(\gamma)\}$ in a finite number of steps are essentially homotopy algorithms. These start with $x^*(\gamma_1 = 0) = y$, and sequentially compute $x^*(\gamma_{i+1})$ from $x^*(\gamma_i)$. One example is the algorithm of [2], that for 1FL has a complexity of $\mathcal{O}(n\log^2 n)$, with a special heap implementation. The best method is the primal path algorithm of [1], with $\mathcal{O}(n\log n)$ complexity.

To understand the complexity of finding $\{x^*(\gamma)\}$, we have to bound T. For 1FL, [26] proves that $T = \mathcal{O}(n)$, and [2], [27] give a different proof of the same fact. These proofs' idea is as follows. The non-differentiable penalty $\sum_t |x_{t+1} - x_t|$ in (2) implies that we have a critical point whenever, as γ increases, for some t, the term $|x_{t+1}^* - x_t^*|$, goes from non-zero to zero, or vice-versa. When this happens, we say that x_{t+1}^* and x_t^* "fuse", or, conversely, that they "un-fuse". One then proves that, as γ increases, variables never "un-fuse". Hence, there are at most n fusing events, and thus $T \leq n$.

Unfortunately, existing proofs do not extend to W1FS. Figure 1 provides an example where variables "fuse" and "un-fuse". Hence, to bound T, we are left with bounds for the generalized lasso which, in a worst case scenario, can be $\Omega(3^n)$ [3].

• Our main contribution is to show that $T = \mathcal{O}(n^2)$, and that, in a worst case scenario, $T = \Omega(n^2)$.

II. Main results

We start by reformulating W1FS, as stated in Theorem 1. In this theorem, and throughout the paper, we use the notation $[n] = \{1, \ldots, n\}$.

Theorem 1. *Let $\tilde{y}_i = -\sum_{t=i}^{n} y_t$ for $i \in [n+1]$ and $\tilde{y}_{n+1} = 0$. Let $\tilde{\alpha}_{i+1} = \alpha_i$ for $i \in [n-1]$ and $\tilde{\alpha}_1 = \tilde{\alpha}_{n+1} = 0$. Let $w^*(\gamma)$ be the unique minimizer of*

$$\min_{w \in \mathbb{R}^n} \sum_{i=1}^{n} (w_{i+1} - w_i)^2 \ s.t. \ |w_i - \tilde{y}_i| \le \gamma \tilde{\alpha}_i, i \in [n+1] \quad (5)$$

We have that $x_t^*(\gamma) = w_{t+1}^*(\gamma) - w_t^*(\gamma)$ for all $t \in [n]$.

See [9], e.g., for a proof of Theorem 1. The Supplementary Material includes another proof, based on the Moreau identity.

Theorem 1 allows us to study the number of linear segments in x^* by studying the number of linear segments in w^*. Indeed, since by Theorem 1 we have that $x_t^*(\gamma) = w_{t+1}^*(\gamma) - w_t^*(\gamma)$, it follows that $x^*(\gamma)$ only changes linear segment if $w^*(\gamma)$ changes linear segment. Hence, if there are at most T different linear segments in $w^*(\gamma)$, there are at most T different linear segments in $x^*(\gamma)$. Similarly, if there are at least T linear segments in $w^*(\gamma)$, and if, for each γ, no two consecutive components of w^* change linear segment, then there are at least T linear segments in $x^*(\gamma)$.

We can make a few simple observations about $w^*(\gamma)$, defined in Theorem 1, which we use to get bounds on T. First we introduce some notation. Figure 2 illustrates its use.

- We refer to w_i^* as the ith variable or ith *point*. We refer to $[\tilde{y}_i - \tilde{\alpha}_i \gamma, \tilde{y}_i + \tilde{\alpha}_i \gamma]$ as the *interval associated to* the ith point. We say that the ith point is *touching its left or right boundary* if $w_i^* = \tilde{y}_i - \tilde{\alpha}_i \gamma$ or if $w_i^* = \tilde{y}_i + \tilde{\alpha}_i \gamma$ respectively. A point that is not touching either side of the boundary of its interval is called *free*. Otherwise, it is called *bounded*.
- We define $F(\gamma) = \{i : |w_i^*(\gamma) - \tilde{y}_i| < \tilde{\alpha}_i \gamma\}$ and $B(\gamma) = [n+1] \backslash F$. In words, a point is in F if and only if it is free. A point is in B if and only if it is bounded.
- We define $s_i(\gamma)$ so that

$$w_i^*(\gamma) = \tilde{y}_i + s_i(\gamma) \tilde{\alpha}_i \gamma. \quad (6)$$

In particular, $s_i(\gamma)$ is 1 (resp. -1) if the point touches its right (resp. left) boundary. For $i = 1$ and $i = n+1$, for which the left and right boundaries are the same, s_i may be chosen arbitrarily.

- For the ith point, we define $i^\triangleleft(\gamma) = \max\{j \in B(\gamma) : j < i\}$ and let $i^\triangleright(\gamma) = \min\{j \in B(\gamma) : i < j\}$. In words, $i^\triangleleft$ and $i^\triangleright$ are the pair of indices of bounded points, smaller and larger than i respectively, that are closest to i.

Whenever clear from the context, we omit the dependency in γ in our expressions. We now list our observations.

1) Since $w^*(\gamma)$ is a continuous function of γ (see e.g. [2]), if a point i has $\tilde{\alpha}_i > 0$, then it cannot touch its left (right) boundary for γ_2 and touch its right (left) boundary for $0 < \gamma_1 < \gamma_2$ without being free for at least one $\gamma \in (\gamma_1, \gamma_2)$. This implies that if B and F do not change for $\gamma \in [\gamma_1, \gamma_2]$, then s_i is constant in $\gamma \in [\gamma_1, \gamma_2]$.
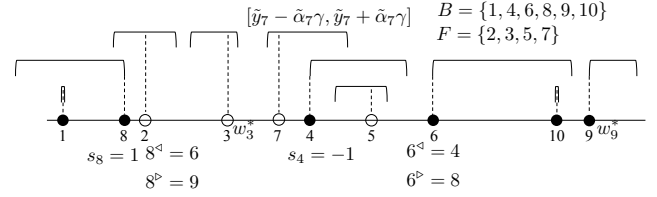

Fig. 2. Illustration of the notation used. Square horizontal brackets represent intervals on the real-line inside which each variable must be. Solid circles represent variables touching their intervals' boundary. B is the set of bounded variables, and F is the set of free variables. $i^\triangleright$ and $i^\triangleleft$ identify closest bounded variables to the left and right of i. In this picture $n = 9$, so $\tilde{\alpha}_1 = \tilde{\alpha}_{n+1} = 0$.

2) If $i \in F$, then w_i^* is determined by the position of the $i^\triangleleft$th and $i^\triangleright$th points, namely,

$$w_i^* = w_{i^\triangleleft}^* f_i + w_{i^\triangleright}^* \bar{f}_i, \quad (7)$$

where $f_i = \frac{i^\triangleright - i}{i^\triangleright - i^\triangleleft}$ and $\bar{f}_i = 1 - f_i$. Equation (7) follows from the fact that, for any $i^\triangleleft < j < i^\triangleright$, w_j^* comes from the solution of the problem $\min_{\{w_j\}} \sum_{j=i^\triangleleft}^{i^\triangleright - 1} (w_{j+1} - w_j)^2$ subject to $w_{i^\triangleleft} = w_{i^\triangleleft}^*$ and $w_{i^\triangleright} = w_{i^\triangleright}^*$. Its solution is that the points $\{w_j^*\}$ must divide the interval $[w_{i^\triangleleft}^*, w_{i^\triangleright}^*]$ in $i^\triangleright - i^\triangleleft$ equal parts, hence (7). Note that, if for all $\gamma \in [\gamma_1, \gamma_2]$, F does not change then, by observation 1, $i^\triangleright$, $i^\triangleleft$, $s_{i^\triangleright}$ and $s_{i^\triangleleft}$ are constant for all $\gamma \in [\gamma_1, \gamma_2]$, and (7) becomes a linear function of γ given by

$$w_i^*(\gamma) = (\tilde{y}_{i^\triangleleft} f_i + \tilde{y}_{i^\triangleright} \bar{f}_i) + \gamma(\tilde{\alpha}_{i^\triangleleft} s_{i^\triangleleft} f_i + \tilde{\alpha}_{i^\triangleright} s_{i^\triangleright} \bar{f}_i). \quad (8)$$

This also implies that a change in linear segment, and hence a critical point, only occurs when F and B change.

3) We give a necessary condition for γ_c to be a critical point, at which w_i^* transitions from being free to bounded as we increase γ. Since w^* is continuous and piecewise linear with a finite number of linear segments (see [2]), we know that F is constant in a small enough interval of the form $I = (\gamma', \gamma_c)$. We can compute $w_i^*(\gamma_c)$ both from (6) and taking the limit of (8) for $\gamma \in I$ and $\gamma \to \gamma_c$ to conclude that for $\gamma = \gamma_c$

$$(\tilde{y}_{i^\triangleleft} f_i + \tilde{y}_{i^\triangleright} \bar{f}_i) + \gamma(\tilde{\alpha}_{i^\triangleleft} s_{i^\triangleleft} f_i + \tilde{\alpha}_{i^\triangleright} s_{i^\triangleright} \bar{f}_i) = \tilde{y}_i + \tilde{\alpha}_i s_i \gamma, \quad (9)$$

where s_i is evaluated at $\gamma = \gamma_c$ and $i^\triangleleft$, $i^\triangleright$, $s_{i^\triangleleft}$ and $s_{i^\triangleright}$ are evaluated at any $\gamma \in I$.

Furthermore, if $s_i(\gamma_c) = +1$, then, for $\gamma \in I$, the left hand side, l.h.s., of (9) is strictly smaller than the right hand side, r.h.s, and, as γ increases to γ_c, the l.h.s must increase until it is equal to the r.h.s. Hence, we conclude that the following relation holds between their rates of growth $\tilde{\alpha}_{i^\triangleleft} s_{i^\triangleleft} f_i + \tilde{\alpha}_{i^\triangleright} s_{i^\triangleright} \bar{f}_i > \tilde{\alpha}_i s_i$. Similarly, if $s_i(\gamma_c) = -1$, then $\tilde{\alpha}_{i^\triangleleft} s_{i^\triangleleft} f_i + \tilde{\alpha}_{i^\triangleright} s_{i^\triangleright} \bar{f}_i < \tilde{\alpha}_i s_i$. We can thus write that, if the ith point transitions from free to bounded, then

$$\tilde{\alpha}_{i^\triangleleft} s_i s_{i^\triangleleft} f_i + \tilde{\alpha}_{i^\triangleright} s_i s_{i^\triangleright} \bar{f}_i > \tilde{\alpha}_i. \quad (10)$$

• Our first result is a new very simple proof, when compared to [2], [26], [27], of the known fact that, for 1FL, $T = \mathcal{O}(n)$. Our second and third result are new altogether.

Theorem 2. *1FL has at most $\mathcal{O}(n)$ different linear segments.*

Proof. It is enough to prove that w^* has at most $\mathcal{O}(n)$ different linear segments. For 1FL, we have $\tilde{\alpha}_{i+1} = 1$ for all $i \in [n-1]$. Therefore, if w_i^* is free and, as γ increases, it becomes bounded, then, by (10), we must have that $1 < s_i s_{i^\triangleleft} f_i +$

$s_i s_{i\triangleright} \bar{f}_i$. This implies that $1 < |s_i s_{i\triangleleft} f_i + s_i s_{i\triangleright} \bar{f}_i| \le f_i + \bar{f}_i = 1$, which is a contradiction. Therefore, a point i never goes from F to B. Since a critical point in $w^*(\gamma)$ only occurs when F changes (by observation 2), and since F can only change by the addition of $n-1$ variables at most, we have $T = \mathcal{O}(n)$. $\square$

Theorem 3. *W1FL has $\mathcal{O}(n^2)$ different linear segments.*

Proof. At any critical value γ_c at which B changes, B might change by multiple points. Some are added, others are removed. Let $i(\gamma_c, r)$ be the rth element changing in B at γ_c, and let $i^\triangleright(\gamma_c, r)$ and $i^\triangleleft(\gamma_c, r)$ be indices of bounded points, larger and smaller than $i(\gamma_c, r)$ respectively, that are closest to $i(\gamma_c, r)$.

Let $i < j$ and consider the set $S_{ij} = \{\gamma_c : (i^\triangleleft(\gamma_c, r), i^\triangleright(\gamma_c, r)) = (i, j)$ for some $r\}$, i.e., the set of critical values at which some point that is being added or removed from B has as its closest bounded points to the left and to the right the points i and j. It suffices to prove that $|S_{ij}| \le 8$, since then total number of critical points is equal to $\sum_{i<j} |S_{ij}| \le 8\binom{n}{2} = \mathcal{O}(n^2)$.

Let $\gamma_c \in S_{ij}$. Let $i < k < j$. By equation (7), we know that $w_k^*(\gamma_c)$ can be expressed as a linear function of γ_c. Hence, together with the fact that we must have $w_k^* \in \tilde{y}_k + \gamma_c[-\tilde{\alpha}_k, \tilde{\alpha}_k]$, we know that $\gamma_c \in [a_k, b_k]$, for some $a_k \ge 0$ and b_k, which depend on s_i and s_j, and the parameters $\tilde{\alpha}, \tilde{y}$ of the problem, where b_k might be ∞. In words, for $w_k^*(\gamma_c)$ to be feasible, γ_c must be in some interval $[a_k, b_k]$.

This implies that $\gamma_c \in \cap_{i<k<j}[a_k, b_k] = [A_{ij}, B_{ij}]$, for some $A_{i,j} \ge 0$ and $B_{i,j}$, which depend on s_i and s_j, and the parameters $\tilde{\alpha}, \tilde{y}$ of the problem, where $B_{i,j}$ might be ∞.

We argue that it must be that γ_c is equal to either A_{ij} or B_{ij}. Indeed, since γ_c is a critical point, if $k = i(\gamma_c, r)$ for some r, then w_k^* must be at the boundary of $\tilde{y}_k + \gamma_c[-\tilde{\alpha}_k, \tilde{\alpha}_k]$. This implies that γ_c is either a_k or b_k, and hence that γ_c is at the boundary of $\cap_{i<k'<j}[a_{k'}, b_{k'}]$. There are 4 choices for the pair (s_i, s_j), and, for each of these choices, γ_c can be either A_{ij} or B_{ij}. Hence, for any pair (i, j), there are at most 8 possible values for γ_c. Thus $|S_{ij}| \le 8$. $\square$

Theorem 4. *There exists α and y such that W1FL has $\Omega(n^2)$ different linear segments. One example is to chose α and y such that $\tilde{\alpha}$ and $\tilde{y}$ satisfy*

$$\tilde{\alpha}_i = (i-1)^2, \forall i \in [n], \tilde{\alpha}_{n+1} = 0, \tag{11}$$

$$\tilde{y}_i = (-1)^i q_i, \forall i \in [n], \tilde{y}_{n+1} = 0, \tag{12}$$

$$q_1 = 1, q_2 = 2, q_{i+2} = 2q_{i+1} - q_i + 2g_{i+2} + 1, \forall i \in [n-2], \tag{13}$$

$$g_3 = 1/3, g_{i+3} = 2g_{i+2} + 1. \tag{14}$$

Remark 5. *This, together with Thrm. 3, implies that there are examples where variables "fuse" $\Theta(n^2)$ times and "un-fuse" $\Theta(n^2)$ times. Furthermore, its proof implies that the difference between the number of "fuse" and "un-fuse" events is $\mathcal{O}(n)$.*

What is the idea behind Theorem (4)? The fact $\tilde{\alpha}_i$ grows super-linearly with i, allows us to have a value of γ around which the ith interval is responsible for driving the behavior of $w^*(\gamma)$. Let us call this the *ith epoch*. The fact that $\tilde{y}_i$ oscillates and diverges exponentially fast with i, drives points between epochs to alternate between being bounded at their right or left

boundary, according to whether $\tilde{y}_i$ is very large and negative or very large and positive. Since there are n epochs, and since, in general, from one epoch to the next, $\Omega(n)$ points change from their right to left boundary (or vice versa), we have $\Omega(n^2)$ "fuse" and "un-fuse" events.

Proof. We are going to produce a set of sufficient conditions that guarantee that the number of critical points in w^* is $\Omega(n^2)$. It is then an algebra exercise, which we omit, to check that these conditions are satisfied by our choice above. Finally, we prove why these same conditions imply that no two consecutive components of w^* change their linear segment at the same time, and hence why x^* also has at least $\Omega(n^2)$ different linear segments.

In particular, our conditions will imply the existence of $\tilde{\alpha}_2, \ldots, \tilde{\alpha}_n \ne 0$ and $\tilde{y}_1, \ldots \tilde{y}_n$, such that there exits a sequence of critical points $0 < \gamma_3 < \gamma_4 < \cdots < \gamma_n$ such that the following two scenarios hold true:

1) for $\gamma = \gamma_{2k}$, $3 \le 2k \le n$, every point w_i^* with $i \in \{2, \ldots, 2k-1\}$ is touching its right boundary;
2) for $\gamma = \gamma_{2k+1}$, $3 \le 2k+1 \le n$, every point w_i^* with $i \in \{2, \ldots, 2k\}$ is touching its left boundary;

Both scenarios imply that, for $\gamma \in [\gamma_r, \gamma_{r+1})$, $r \in \{3, \ldots, n-1\}$, every point w_i^*, $i \in \{2, \ldots, r-1\}$, changes from touching one side of its boundary to the other side of its boundary. Note that since $\tilde{\alpha}_2, \ldots \tilde{\alpha}_n \ne 0$, if $\gamma > 0$, a point cannot be simultaneously touching its left and right boundary. Hence, $w^*(\gamma)$ has at least $r - 2$ critical points in $\gamma \in [\gamma_r, \gamma_{r+1})$. Hence, for $\gamma \in [\gamma_2, \gamma_n]$, there are at least

$$1 + \cdots + n - 1 = n(n-1)/2 \tag{15}$$

critical points.

Let us be more specific about the two scenarios. In Scenario 1, in addition to what we have already described, for $\gamma = \gamma_r$ and $r = 2k$, we also want the intervals $[\tilde{y}_1 - \tilde{\alpha}_1\gamma, \tilde{y}_1 + \tilde{\alpha}_1\gamma]$, ..., $[\tilde{y}_{r-1} - \tilde{\alpha}_{r-1}\gamma, \tilde{y}_{r-1} + \tilde{\alpha}_{r-1}\gamma]$ to be nested, with the intervals for large i containing the intervals for small i. We also want the left boundary of the interval $[\tilde{y}_r - \tilde{\alpha}_r\gamma, \tilde{y}_r + \tilde{\alpha}_r\gamma]$ to be larger than the right most limit of all the intervals involving indices smaller than r. The following picture illustrates these conditions.
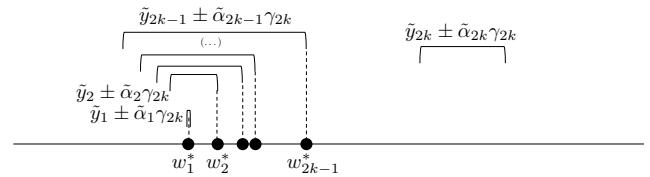
In Scenario 2, in addition to what we have already described, for $\gamma = \gamma_r$ and $r = 2k + 1$, we want the same conditions as in Scenario 1 but now with left and right reversed. The following picture illustrates these conditions.
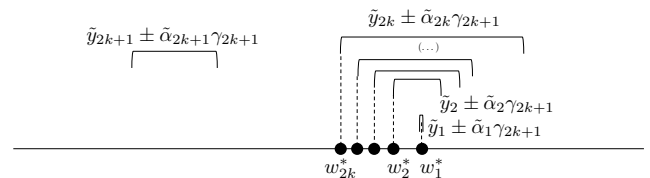
Note that in both scenarios, we do not care where the points $r+1, \ldots, n$ are.

As we explain next, a set of *sufficient* conditions for Scenario 1 (Scenario 2) to hold is that, for all $r \in \{3, \ldots, n\}$,

$$|\tilde{y}_{i+1} - \tilde{y}_i| < (\tilde{\alpha}_{i+1} - \tilde{\alpha}_i)\gamma_r, \forall i \in [r-2], \quad (16)$$

$$|\tilde{y}_{i+2} - 2\tilde{y}_{i+1} + \tilde{y}_i| < (\tilde{\alpha}_{i+2} - 2\tilde{\alpha}_{i+1} + \tilde{\alpha}_i)\gamma_r, \forall i \in [r-3], \quad (17)$$

$$(-1)^r(\tilde{y}_r - 2\tilde{y}_{r-1} + \tilde{y}_{r-2}) > (\tilde{\alpha}_r + 2\tilde{\alpha}_{r-1} - \tilde{\alpha}_{r-2})\gamma_r. \quad (18)$$

Condition (16) directly implies that the first $r-1$ intervals are nested. Condition (17) implies that if the $(i+2)$th and ith points are touching their right (left) boundary, then the $(i+1)$th point must be touching its right (left) boundary. This can be seen by solving the simple quadratic problem $\min_w (w_{i+2} - w_{i+1})^2 + (w_{i+1} - w_i)^2$ subject to w_{i+2} and w_{i+1} behind at their boundaries, and w_{i+1} being inside its interval. Condition (18) implies that, no matter where w_r^* is, even if w_r^* is touching its left (right) boundary, the $(r-1)$th point must touch its right (left) boundary. This can also be seen by solving a simple quadratic problem involving three variables. These three conditions together imply that the first $r-1$ points are touching their right (left) boundary.

As we explain next, these conditions are in turn implied by the following set of conditions. These also imply that γ_r is increasing.

$$\tilde{\alpha}_{i+2} - 2\tilde{\alpha}_{i+1} + \tilde{\alpha}_i > 0, \forall i \in [n-2], \quad (19)$$

$$\tilde{\alpha}_{i+1} - \tilde{\alpha}_i > 0, \forall i \in [n-1], \quad (20)$$

$$\gamma_3 > \frac{q_2 - q_1}{\tilde{\alpha}_2 - \tilde{\alpha}_1}, \quad (21)$$

$$\gamma_{i+3} > \max\left\{\gamma_{i+2}, \frac{q_{i+2} - 2q_{i+1} + q_i}{\tilde{\alpha}_{i+2} - 2\tilde{\alpha}_{i+1} + \tilde{\alpha}_i}, \frac{q_{i+2} - q_{i+1}}{\tilde{\alpha}_{i+2} - \tilde{\alpha}_{i+1}}\right\}, \quad (22)$$
$$\forall i \in [n-3],$$

$$q_2 > q_1, \quad (23)$$

$$q_{i+2} > 2q_{i+1} - q_i + (\tilde{\alpha}_{i+2} + 2\tilde{\alpha}_{i+1} - \tilde{\alpha}_i)\gamma_{i+2}, \forall i \in [n-2], \quad (24)$$

$$\tilde{y}_i = (-1)^i q_i, \forall i \in [n]. \quad (25)$$

Conditions (20), (21) and (23) imply that $\gamma_3 > 0$. Condition (22) further implies that $0 < \gamma_3 < \cdots < \gamma_n$. Condition (24) and (25) imply condition (18). Since $\tilde{\alpha} > 0$, condition (19) implies that $\tilde{\alpha}_{i+2} + 2\tilde{\alpha}_{i+1} - \tilde{\alpha}_i > 0$ for all $i \in [n-2]$ and thus, since $\gamma_{i+2} > 0$, condition (23) together with (24) imply that $q_{i+2} - 2q_{i+1} + q_i > 0$ for all $i \in [n-2]$ and that $q_{i+1} - q_i > 0$ for all $i \in [n-1]$. Therefore, $|\tilde{y}_{i+2} - 2\tilde{y}_{i+1} + \tilde{y}_i| = q_{i+2} - 2q_{i+1} + q_i$ for all $i \in [n-2]$ and $|\tilde{y}_{i+1} - \tilde{y}_i| = q_{i+1} - q_i$ for all $i \in [n-1]$. These two equations, together with conditions (19), (20), (21) and (22) imply that (16) holds for all $i \in [n-1]$, if we replace γ_r by γ_{i+2}. These also imply that (17) holds for all $i \in [n-2]$, if we replace γ_r by γ_{i+3} there. But since γ_i is increasing, we have that (16) and (17) hold true as specified.

We finally argue by contradiction why conditions (19)-(25) also imply that no two consecutive components of w^* stop touching, or start touching, their boundary at the same γ. We prove this only for values of γ that are in some of the intervals $[\gamma_r, \gamma_{r+1})$ that contribute to our n^2 estimate in (15).

Assume that both w_i^* and w_{i+1}^* have a critical point at $\gamma = \gamma_c$. Assume also that $\gamma_c \in [\gamma_r, \gamma_{r+1})$ for some $r > i+1$. Note that γ_c must satisfy this condition if it is to contribute

for our n^2 estimate in (15). It must be that $|w_{i-1}^* - w_i^*| = |w_i^* - w_{i+1}^*| = |w_{i+1}^* - w_{i+2}^*|$. At the same time, and as we saw in the previous paragraph, conditions (19)-(25) imply that $|\tilde{y}_{i+1} - 2\tilde{y}_i + \tilde{y}_{i-1}| < (\tilde{\alpha}_{i+1} - 2\tilde{\alpha}_i + \tilde{\alpha}_{i-1})\gamma_r \le (\tilde{\alpha}_{i+1} - 2\tilde{\alpha}_i + \tilde{\alpha}_{i-1})\gamma_c$, which in turn implies that $|w_i^* - w_{i-1}^*| < |w_{i+1}^* - w_i^*|$, which is a contradiction. $\square$

III. NUMERICAL EXPERIMENTS

Figure 3-(left) shows a numerical computation of the number of critical points as a function of n for the example (11)-(14). As Theorem 4 predicts, the number of critical points grows quadratically with n. One can also observe that difference between "fuse" and "un-fuse" events is $\mathcal{O}(n)$. For Figure 3-(right), we generated 100 random sets of α and y, and, for each size n, we show on the y-axis the average number of "fuse" events and "un-fuse" events observed over these 100 runs. Each α_i was sampled from a uniform distribution in $[0, 1]$, independently across α's, and each y_i was sampled from a $\mathcal{N}(0, \sqrt{10})$, independently across y's. Although Theorem 4 tells us that we can observe $\Omega(n^2)$ "fuse" and "un-fuse" events, in our random instances, "un-fuse" events are rare and both types of events seem to grow linearly with n.

Critical points can be computed using, for example, [2]. In Supplementary Material we give simple algorithm which we use to compute the critical points based on Theorem 1.
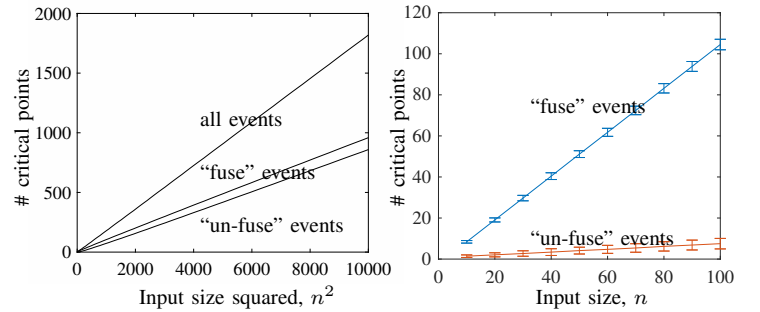

Fig. 3. (Left) Number of critical points as function of n^2 for the example of Theorem 4; (Right) Number of "fuse" and "un-fuse" events in random instances of W1FL, as a function of n.

IV. FUTURE WORK

The weighted fused lasso on an input of size n is substantially different from the equal-weights fused lasso: two consecutive components can both become equal ("fuse") and become unequal ("un-fuse") multiple times along the solution path. We have shown that there are instances with $\Omega(n^2)$ of these "fuse"/"un-fuse" events, and that no instance can have more than $\mathcal{O}(n^2)$ events. We also gave a very simple proof of why, in the equal-weights fused lasso, there are $\mathcal{O}(n)$ events.

Future work should include finding conditions for the weights α and input y under which a $\mathcal{O}(n)$ bound holds, and conditions for which the number of "fuse" and "un-fuse" events are substantially different. It would then be useful to compute how likely it is for these conditions to be satisfied, under different stochastic models for the input.

ACKNOWLEDGMENT

References

[1] H. Hoefling, "A path algorithm for the fused lasso signal approximator," *Journal of Computational and Graphical Statistics*, vol. 19, no. 4, pp. 984–1006, 2010.

[2] R. J. Tibshirani, *The solution path of the generalized lasso*. Stanford University, 2011.

[3] J. Mairal and B. Yu, "Complexity analysis of the lasso regularization path," *arXiv preprint arXiv:1205.0079*, 2012.

[4] A. Chambolle, V. Duval, G. Peyré, and C. Poon, "Geometric properties of solutions to the total variation denoising problem," *Inverse Problems*, vol. 33, no. 1, p. 015002, 2016.

[5] P. L. Davies and A. Kovac, "Local extremes, runs, strings and multiresolution," *Annals of Statistics*, pp. 1–48, 2001.

[6] N. A. Johnson, "A dynamic programming algorithm for the fused lasso and l 0-segmentation," *Journal of Computational and Graphical Statistics*, vol. 22, no. 2, pp. 246–260, 2013.

[7] L. Condat, "A direct algorithm for 1-d total variation denoising," *IEEE Signal Processing Letters*, vol. 20, no. 11, pp. 1054–1057, 2013.

[8] "Total variation denoising of 1-d signals c code." http://www.gipsa-lab. grenoble-inp.fr/~laurent.condat/download/Condat_TV_1D_v2.c. Accessed: 2017-12-10.

[9] A. Barbero and S. Sra, "Modular proximal optimization for multidimensional total-variation regularization," *arXiv preprint arXiv:1411.0589*, 2014.

[10] L. Dümbgen, A. Kovac, *et al.*, "Extensions of smoothing via taut strings," *Electronic Journal of Statistics*, vol. 3, pp. 41–75, 2009.

[11] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2419–2434, 2009.

[12] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of mathematical imaging and vision*, vol. 40, no. 1, pp. 120–145, 2011.

[13] B. Wahlberg, S. Boyd, M. Annergren, and Y. Wang, "An admm algorithm for a class of total variation regularized estimation problems," *IFAC Proceedings Volumes*, vol. 45, no. 16, pp. 83–88, 2012.

[14] S. Bonettini and V. Ruggiero, "On the convergence of primal–dual hybrid gradient algorithms for total variation image restoration," *Journal of Mathematical Imaging and Vision*, vol. 44, no. 3, pp. 236–253, 2012.

[15] G.-B. Ye and X. Xie, "Split bregman method for large scale fused lasso," *Computational Statistics & Data Analysis*, vol. 55, no. 4, pp. 1552–1569, 2011.

[16] A. Barbero and S. Sra, "Fast newton-type methods for total variation regularization," in *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pp. 313–320, Citeseer, 2011.

[17] L. Condat, "A primal–dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms," *Journal of Optimization Theory and Applications*, vol. 158, no. 2, pp. 460–479, 2013.

[18] A. Ramdas and R. J. Tibshirani, "Fast and flexible admm algorithms for trend filtering," *Journal of Computational and Graphical Statistics*, vol. 25, no. 3, pp. 839–858, 2016.

[19] Z.-F. Pang and Y. Duan, "Primal-dual method to the minimized surface regularization for image restoration," *arXiv preprint arXiv:1605.09113*, 2016.

[20] G. França and J. Bento, "An explicit rate bound for over-relaxed admm," in *Information Theory (ISIT), 2016 IEEE International Symposium on*, pp. 2104–2108, IEEE, 2016.

[21] G. França and J. Bento, "How is distributed admm affected by network topology?," *arXiv preprint arXiv:1710.00889*, 2017.

[22] R. J. Tibshirani *et al.*, "The lasso problem and uniqueness," *Electronic Journal of Statistics*, vol. 7, pp. 1456–1490, 2013.

[23] Y. Dong, M. Hintermüller, and M. M. Rincon-Camacho, "Automated regularization parameter selection in multi-scale total variation models for image restoration," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 82–104, 2011.

[24] Y.-W. Wen and R. H. Chan, "Parameter selection for total-variation-based image restoration using discrepancy principle," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1770–1781, 2012.

[25] I. Loris, "On the performance of algorithms for the minimization of l1-penalized functionals," *Inverse Problems*, vol. 25, no. 3, p. 035008, 2009.

[26] J. Friedman, T. Hastie, H. Höfling, R. Tibshirani, *et al.*, "Pathwise coordinate optimization," *The Annals of Applied Statistics*, vol. 1, no. 2, pp. 302–332, 2007.

[27] R. J. Tibshirani and J. Taylor, "Supplement: Proofs and technical details for the solution path of the generalized lasso,"

[28] P. R. Halmos, *Measure theory*, vol. 18. Springer, 2013.

[29] R. K. Sundaram, *A first course in optimization theory*. Cambridge university press, 1996.

Supplementary Material to "On the Complexity of the Weighted fused Lasso"

V. PROOF OF THEOREM 1

We begin by reviewing Moreau's identity. Let $f(x)$ be a closed proper convex function and let

$$F(y) = \arg\min_x f(x) + \frac{1}{2}\|x - y\|_2^2 \qquad (26)$$

be its proximal operator. Let

$$\hat{f}(x) = \sup_z \langle z, x\rangle - f(z) \qquad (27)$$

be the Fenchel dual of f and let

$$\hat{F}(y) = \arg\min_x \hat{f}(x) + \frac{1}{2}\|x - y\|_2^2 \qquad (28)$$

be its proximal operator. Moreau's identity is

$$F(y) = y - \hat{F}(y). \qquad (29)$$

Proof of Theorem 1. If $f(x) = \sum_{t=1}^{n-1} \gamma\alpha_t|x_{t+1} - x_t|$, then $x^*(\gamma) = F(y)$. Therefore, Theorem 1 amounts to an optimization problem to compute $F^*(y)$, from which we can compute F, and hence x^*, using (29).

We start by making a change of variables in

$$f^*(x) = \sup_z \sum_{t=1}^{n} z_t x_t - \sum_{t=1}^{n-1} \alpha_t|z_{t+1} - z_t|. \qquad (30)$$

Let $h_0 = z_1$ and $h_t = z_{t+1} - z_t$, for $t \in [n-1]$. This implies that $z_{t+1} = \sum_{i=0}^{t} h_i$ and thus that $\sum_{t=1}^{n} z_t x_t = \sum_{t=1}^{n} \sum_{i=0}^{t-1} h_i x_t = \sum_{t=0}^{n-1} h_t(\sum_{i=t+1}^{n} x_i) = \sum_{t=0}^{n-1} h_t u_t$, where we have defined $u_t = \sum_{i=t+1}^{n} x_i$ Therefore, we can rewrite (30) as

$$f^*(x) = \sup_h \sum_{t=0}^{n-1} h_t u_t - \sum_{t=0}^{n-1} \gamma\alpha_t|h_t|, \qquad (31)$$

where we have extended α such that $\alpha_0 = 0$. Problem (31) breaks down into n independent one-dimensional problems of the form

$$\sup_{h_t} h_t u_t - \gamma\alpha_t|h_t|, \qquad (32)$$

that have solution 0 if $u_t \in [-\gamma\alpha_t, \gamma\alpha_t]$, and ∞ otherwise. Therefore, $\hat{f}(x) = \infty$ if $u_t = \sum_{i=t+1}^{n} x_i \notin [-\gamma\alpha_t, \gamma\alpha_t]$ for some $t = 0, \cdots, n-1$, and $\hat{f}(x) = 0$ otherwise.

We can now write

$$F(y) = y - \hat{F}(y) = y - \arg\min_x \sum_{t=1}^{n}(x_t - y_t)^2 \qquad (33)$$

subject to $\sum_{i=t+1}^{n} x_i \in [-\gamma\alpha_t, \gamma\alpha_t], t = 0, \cdots, n-1.$

Finally, we make the following change of variable in (33): $w_t = \sum_{i=t}^{n}(x_i - y_i)$ for all $t \in [n+1]$, where we define $w_{n+1} = 0$. This implies that $x_i = y_i - w_{i+1} + w_i$ for $i \in [n]$, and hence that

$$x_i^* = F_i(y) = y_i - \hat{F}_i(y) = y_i - (y_i - w_{i+1}^* + w_i^*) \qquad (34)$$

$$= w_{i+1}^* - w_i^* \qquad (35)$$

for all $i \in [n]$, where

$$w^* = \arg\min_w \sum_{t=1}^{n}(w_{t+1} - w_t)^2 \qquad (36)$$

subject to $w_{t+1} - \sum_{i=t+1}^{n} y_i \in [-\gamma\alpha_t, \gamma\alpha_t], t = 0, \cdots, n-1.$

$\square$

VI. SIMPLE ALGORITHM TO COMPUTE CRITICAL POINTS

Before we introduce the algorithm, we need a fourth observation, in addition to the three observations already made in the main text. For this observation to be valid, we assume that the α's are not in some set of measure zero on the space of all possible α's. The following technical results, whose proofs are standard, e.g. [28], [29], will be used to extend these observations to any set of α's. We include a self contained proof of these lemmas in Section VII.

Lemma 6. *Let $x \in X \subset \mathbb{R}^k$, where X has zero Lebesgue measure. For any $\epsilon > 0$, there exists a point $y \notin X$ such that $\|x - y\| < \epsilon$.*

Lemma 7. *The function $x^*(\gamma, \alpha)$ is continuous at every point of the domain $(\gamma, \alpha) \geq 0$.*

Remark 8. *Since $w_{t+1}^*(\gamma, \alpha) = \tilde{y}_1 + \sum_{i=1}^{t} x_i^*(\gamma, \alpha)$, this also proves that $w^*(\gamma, \alpha)$ is continuous at every point of its domain.*

4) We can assume, without loss of generality, that at most one component of w^* transitions from free to bounded, or vice versa, at each γ. This follows from the fact that, for two indices i and j (or more) to satisfy (9) for the same γ_c, the values of $\tilde{\alpha}_i$, $\tilde{\alpha}_j$, $\tilde{\alpha}_{i\triangleright}$, $\tilde{\alpha}_{i\triangleleft}$, $\tilde{\alpha}_{j\triangleright}$ and $\tilde{\alpha}_{i\triangleleft}$ must belong to some set ζ of measure zero (in the space of possible α's). Using Lemma 6 and Lemma 7, we can then extend our arguments made outside ζ to this set as well.

Our four observations allows us to describe a simple direct algorithm to compute the path $w^*(\gamma)$. Its interpretation is simple: start with $\gamma = 0$. Increase γ and, as the intervals $\tilde{y}_i + \gamma[-\tilde{\alpha}_i, \tilde{\alpha}_i]$ grow larger, keep track of which points are touching either limit of its interval. For each interval of values of γ for which B is fixed, we can use (8) to compute how each point moves, figure out which next point that will become free, or bounded, and hence compute the next B.

The following algorithm describes how to compute $\{\gamma_i\}_{i=1}^{T-1}$, the critical points. However, it can be modified to compute $\{w^*(\gamma_i)\}_{i=1}^{T-1}$ with only a multiplying factor slow down in the run time.

1) Start with $\gamma_1 = 0$, $F = \{\}$ and iteration number $r = 1$. All points are bounded.
2) At iteration r, use (9) to find γ's at which an $i \notin F$ *might* enter F or at which an $i \in F$ *might* leave F. Store these values as $\gamma_c^{(1)}, \gamma_c^{(2)}, \ldots$ and $\gamma_c'^{(1)}, \gamma_c'^{(2)}, \ldots$ respectively, where the upper index (i) in $\gamma_c^{(i)}$ and $\gamma_c'^{(i)}$ refers to the index of the point that might enter or leave F. If point i cannot leave F in round r, then $\gamma_c^{(i)} = \infty$. If point i cannot enter F in round r, then $\gamma_c'^{(i)} = \infty$.

These values are not all critical points on $w^*(\gamma)$, but all critical points satisfy (9), and, in particular, the smallest value produced corresponds to the next critical point.

3) Append or remove from F the index

$$i_c = \min\{\min_{i:\gamma_c^{(i)}>\gamma_r}\gamma_c^{(i)}, \min_{i:\gamma_c'^{(i)}>\gamma_r}\gamma_c'^{(i)}\}. \qquad (37)$$

Set $\gamma_{r+1} = \gamma_c^{(i_c)}$ or $\gamma_{r+1} = \gamma_c'^{(i_c)}$. Set $\gamma_c^{(i)} = \infty$ or $\gamma_c'^{(i)} = \infty$, the choice depending on whether there is a point leaving, or entering F. We do this last assignment to avoid a cycle of removing and adding the same point to F, ad infinitum, for the same value of γ, without the algorithm progressing.

4) Update $\gamma_c^{(j)}$ or $\gamma_c'^{(j)}$ for all $j \in \{i^\lhd, \ldots, i-1, i+1, \ldots, i^\rhd\}$. All the other previously computed $\gamma_c^{(j)}$ and $\gamma_c'^{(j)}$ will be the same. For later purposes, let us call by k_r the number of values updated in this step.

5) Terminate if $B = \{1, n+1\}$, otherwise go back to step 2.

In the above algorithm, it is convenient to represent B as a linked list such that we can access its elements in the order of their indices. We do not explicitly represent F. An element is in F if it is not in the linked list B. This representation for B also allows us to loop over the elements in F in order of their indices, by looping over the elements of B and, for any two consecutive elements in B, say $a < b$, looping over all indices $a+1, \ldots, b-1$, which we know must be in F. Given a point i, either in B or in F, this allows us to determine, in $\mathcal{O}(1)$ steps, the indices $i^\lhd$ and $i^\rhd$. We can also add and remove points from B in $\mathcal{O}(1)$ steps while keeping the list ordered.

If we use a binary minimum heap to keep track of the minimum value of across $\{\gamma_c^{(j)}\}$ and $\{\gamma_c'^{(j)}\}$, we pay a computational cost of $\mathcal{O}(\log n)$ each time we update $\gamma_c^{(j)}$ or $\gamma_c'^{(j)}$ for some j. Therefore, the complexity of this algorithm is

$$\mathcal{O}\left(\left(\sum_{r=1}^{T}k_r\right)\log n\right) \qquad (38)$$

where, we recall, T is the number of critical points in the path $\{w^*(\gamma)\}$, and n is the number of components in the input y. In the particular case of 1FL, where $\alpha_i = 1 \; \forall i$, we have $T \leq n$ (See e.g. Theorem 2) and points only enter F, they never leave F. Therefore, we only need to use (9) for points in B and thus $k_r = 2 \; \forall r$. This leads us to $\mathcal{O}(n \log n)$, just like in [1].

VII. PROOF OF TECHNICAL LEMMA 6 AND LEMMA 7

Proof of Lemma 6. The proof follows by contradiction. Assume that there exists $\epsilon > 0$ such that, for all y with $\|x - y\| < \epsilon$, we have $y \in X$. X contains a ball of size $\epsilon/2$, and thus has non-zero measure. □

Proof of Lemma 7. Since $(\gamma\alpha_1, \ldots, \gamma\alpha_{n+1})$ is continuous as a function of (γ, α), it follows that, to prove that x^* is continuous as a function of (γ, α), we only need to prove that x^* is continuous as a function of α.

We now assume γ fixed, and for simplicity write $x^*(\gamma)$ as x^* or $x^*(\alpha)$. The same goes for all the variables introduced below. The proof proceeds in two steps. First we show that x^* can be obtained from a linear transformation of a point z^*

defined as the point in the convex polytope $\mathcal{P}(\alpha) = \{z \in \mathbb{R}^n : |\tilde{y}_i - \sum_{r=1}^{i} z_r| \leq \gamma\tilde{\alpha}_{i-1} \text{ for all } i \in [n+1]\}$ that is closest to the origin. Second we show that z^* is continuous in α.

To establish the first step, just make the change of variable $z_1 = w_1$, $z_{i+1} = w_{i+1} - w_i$, $i > 1$ in (5).

To establish the second step, we first make three observations. (Obs. 1) If $|\alpha' - \alpha| < \delta$, then for any $z' \in \mathcal{P}(\alpha')$, there exists $z \in \mathcal{P}(\alpha)$ such that $\|z - z'\| < \epsilon_1(\delta)$, where $\epsilon_1(\delta)$ converges to zero as δ converges to zero. (Obs. 2) If $|\alpha' - \alpha| < \delta$, then $\|z^*(\alpha')\| < \|z^*(\alpha)\| + \epsilon_2(\delta)$, where $\epsilon_2(\delta)$ converges to zero as δ converges to zero. (Obs. 3) If $z \in \mathcal{P}(\alpha)$ is such that $\|z\| < \|z^*\| + \delta$, then $\|z - z^*\| < \epsilon_3(\delta)$, where $\epsilon_3(\delta)$ converges to zero as δ converges to zero.

Obs. 1 follows because the faces of the polytope change continuously with α. Obs. 2 follows from Obs. 1, since, if $z \in \mathcal{P}(\alpha')$ is the closet point to $z^*(\alpha)$, then, by Obs. 1, $\|z^*(\alpha)\| > \|z\| - \epsilon_1(\delta) \geq \|z^*(\alpha')\| - \epsilon_1(\delta)$. Obs. 3 is trivial when the origin is in the interior of $\mathcal{P}(\alpha)$, so we focus on the other case. Let $d = \sup \|z - z^*\|$ be such that $z \in \mathcal{P}$, and $\|z\| < \|z^*\| + \delta$, and let $d' = \sup \|z - z^*\|$ be such that $z \in \mathcal{H}$, and $\|z\| < \|z^*\| + \delta$, where $\mathcal{H}$ is the half-plane defined by $z^\top z^* \geq \|z^*\|^2$. Since $\mathcal{P}(\alpha)$ is convex, $\mathcal{P}(\alpha) \subseteq \mathcal{H}$ and thus $d \leq d' < \sqrt{(\|z^*\| + \delta)^2 - \|z^*\|^2} = \sqrt{\delta}\sqrt{2\|z^*\| + \delta}$, where the last inequality follows by simple geometry. Therefore, $\|z - z^*\| < \sqrt{\delta}\sqrt{2\|z^*\| + \delta}$, which converges to zero as δ converges to zero.

Now let $|\alpha' - \alpha| < \delta$ and let $q \in \mathcal{P}(\alpha)$ be the closet point to $z^*(\alpha')$. By Obs. 1, $\|z^*(\alpha') - q\| < \epsilon_1(\delta)$ and thus $\|q\| < \|z^*(\alpha')\| + \epsilon_1(\delta)$. By Obs. 2, $\|q\| < \|z^*(\alpha)\| + \epsilon_2(\delta) + \epsilon_1(\delta)$. By Obs. 3, $\|q - z^*(\alpha)\| < \epsilon_3(\epsilon_2(\delta) + \epsilon_1(\delta))$. We can finally write, $\|z^*(\alpha') - z^*(\alpha)\| = \|z^*(\alpha') - q + q - z^*(\alpha)\| \leq \|z^*(\alpha') - q\| + \|q - z^*(\alpha)\| \leq \epsilon_1(\delta) + \epsilon_3(\epsilon_2(\delta) + \epsilon_1(\delta))$. This proves continuity since the right hand side converges to zero as δ converges to zero.

□