On the Complexity of the Weighted Fused Lasso

José Bento jose.bento@bc.edu Surjyendu Ray raysc@bc.edu

***Abstract*—The solution path of the 1D fused lasso for an n-dimensional input is piecewise linear with $\mathcal{O}(n)$ 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 $\Omega(3^n)$ segments [3]. In this paper, we prove that the number of segments in the solution path of the weighted fused lasso is $\mathcal{O}(n^3)$, and for some instances $\Omega(n^2)$. 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

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

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

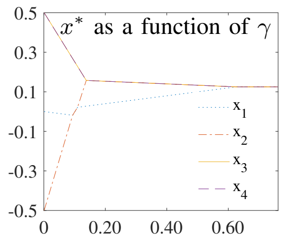
$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\ \frac{1}{2}\sum_{t=1}^{n}(x_t-y_t)^2+\gamma\sum_{t=1}^{n-1}\alpha_t|x_{t+1}-x_t|. \tag{2}$$

We called this the *weighted 1-D fused lasso* (W1FL) with input y and weights $\alpha_t\geq 0\ \forall t$, and 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*, which solve 1FL exactly in a finite number of steps, include the Taut String algorithm, [5], with a worst case complexity of $\mathcal{O}(n^2)$; the algorithm of [6], based on dynamic programing, with $\mathcal{O}(n)$ complexity even in the worst case; 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 a version, [8], that finishes in $\mathcal{O}(n)$ steps. 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 other methods, [20], [21]. However, in many applications, when precision is crucial, or when implementing a termination procedure has a non-negligible computational cost, direct algorithm are preferred.

Frequently, we are not just interested in solving (2) for a single γ. Let $x^*(\gamma)$ be the unique solution of (2) [1]. 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 best the result in a given application, [23], [24].

[1]The generalized lasso (1) might not always have a unique solution [22].



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.

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

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

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

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

and 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) are equivalent problems.

Efficiently characterizing $\{x^*(\gamma)\}$ is possible in part 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* $\gamma_1,\gamma_2,...,\gamma_T$ at which $x^*(\gamma)$ changes from one linear segment to another, and the value of $x^*(\gamma)$ at these γs.

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 latter 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, some 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". Then one 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 I is 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 reduce the gap between the upper estimate and worst case estimate for T.

II. Main results

We start by reformulating W1FS, as stated in Theorem 1. In this theorem, and throughout the paper, we use the notation $[n] = \{1, \ldots, n\}$. This theorem says that, for each γ, we can look at problem (2) as finding the equilibrium state of a physical system of identical springs connected in a one-dimensional chain, where the right extremity of the ith spring is confined to $[\tilde{y}_i - \gamma\tilde{\alpha}_i, \tilde{y}_i + \gamma\tilde{\alpha}_i]$.

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

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

We have that $x^*(\gamma)_t = w^*(\gamma)_{t+1} - w^*(\gamma)_t$ *for all* $t \in [n]$*.*

Note that (5) is related to the interpretation made in the Taut String algorithm [5], which further shows that (5) can be studied via the objective $\sum_{i=1}^{n} \sqrt{1 + (w_{i+1} - w_i)^2}$. This objective is interpreted as a sequence of connected strings, not springs, whose ends are constrained. See [9], for example, for a proof of Theorem 1. We include a self contained proof, based on the Moreau identity, as Supplementary Material.

Theorem 1 allows us to study the number of different linear segments in x^* by studying the number of different linear segments in w^*. Indeed, since by Theorem 1 we have $x^*(\gamma)_t = w^*(\gamma)_{t+1} - w^*(\gamma)_t$, 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 different 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 different linear segments in $x^*(\gamma)$.

From the representation in Theorem 1, we can make a few straightforward observations about $w^*(\gamma)$ that we use to prove our bounds on T.

We begin by introducing some useful nomenclature. 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 *non-free*.
- We define $F(\gamma) = \{i : |w^*(\gamma)_i - \tilde{y}_i| < \tilde{\alpha}_i\gamma\}$ and $B(\gamma) = [n+1] \backslash F$. In words, a point is in F if and only if it is it is free. A point is in B if and only if it is not free.
- We define $s_i(\gamma) = 1$ if and only if $w^*(\gamma)_i = \tilde{y}_i + \tilde{\alpha}_i\gamma$ and $s_i(\gamma) = -1$ if and only if $w^*(\gamma)_i = \tilde{y}_i - \tilde{\alpha}_i\gamma$. For $i = 1$ and $i = n+1$, for which the left and right boundaries are the same, we choose s_i by convention. It does not matter which values we choose.
- 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 non-free points, smaller and larger than i respectively, that are closer to i.

For simplicity, and whenever clear from the context, we omit the dependency in γ in our expressions.

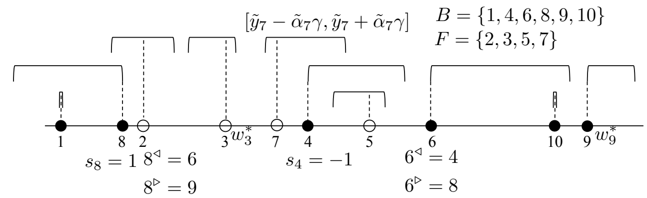


Fig. 2. Illustration of the notation used. Square horizontal brackets represent intervals inside which each variable must be. Solid circles represent variables touching their intervals' boundary. In this picture $n = 9$, so $\tilde{\alpha}_1 = \tilde{\alpha}_{n+1} = 0$.

The first observation is that, 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 all $\gamma \in [\gamma_1, \gamma_2]$, then s_i is constant in $\gamma \in [\gamma_1, \gamma_2]$.

The second observation is that if $i \in F$, then w_i^* is dictated by the position of the $i^{\triangleleft}$th and $i^{\triangleright}$th points, namely,

$$w_i^* = w_{i^{\triangleleft}}^* f_i + w_{i^{\triangleright}}^* \bar{f}_i, \quad (6)$$

where $f_i = \frac{i^{\triangleright} - i}{i^{\triangleright} - i^{\triangleleft}}$ and $\bar{f}_i = 1 - f_i$. Equation (6) follows from the fact that, since all the springs' extremities between $i^{\triangleleft}$ and $i^{\triangleright}$ are free by the definition of $i^{\triangleleft}$ and $i^{\triangleright}$, the springs must be equally distended and thus divide the interval $[w^*(\gamma)_{i^{\triangleleft}}, w^*(\gamma)_{i^{\triangleright}}]$ in $i^{\triangleright} - i^{\triangleleft}$ equal parts. Note that, if for all $\gamma \in [\gamma_1, \gamma_2]$, F does not change then, by the first observation, $i^{\triangleright}$, $i^{\triangleleft}$, $s_{i^{\triangleright}}$ and $s_{i^{\triangleleft}}$ are constant for all $\gamma \in [\gamma_1, \gamma_2]$ and (6) becomes a linear function of γ given by

$$w_i^* = (\tilde{y}_{i^{\triangleleft}} f_i + \tilde{y}_{i^{\triangleright}} \bar{f}_i) + \gamma(\tilde{\alpha}_{i^{\triangleleft}} s_{i^{\triangleleft}} f_i + \tilde{\alpha}_{i^{\triangleright}} s_{i^{\triangleright}} \bar{f}_i). \quad (7)$$

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

The third observation is a necessary condition for any critical point γ_c at which w_i^* transitions from being free to non-free, 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 then use (7) for all $\gamma \in I$ and the continuity property to conclude that γ_c must satisfy

$$(\tilde{y}_{i^{\triangleleft}} f_i + \tilde{y}_{i^{\triangleright}} \bar{f}_i) + \gamma_c(\tilde{\alpha}_{i^{\triangleleft}} s_{i^{\triangleleft}} f_i + \tilde{\alpha}_{i^{\triangleright}} s_{i^{\triangleright}} \bar{f}_i) = \tilde{y}_i + \tilde{\alpha}_i s_i \gamma_c, \quad (8)$$

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 point in I.

Furthermore, assume that $s_i(\gamma_c) = +1$, then, for $\gamma \in I$, the left hand side, l.h.s., of (8) 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 non-free, then

$$\tilde{\alpha}_{i^{\triangleleft}} s_i s_{i^{\triangleleft}} f_i + \tilde{\alpha}_{i^{\triangleright}} s_i s_{i^{\triangleright}} \bar{f}_i > \tilde{\alpha}_i. \quad (9)$$

The fourth observation is that we can also characterize γ_c at which, as we increase γ, w_i^* transitions from being non-free to free. We simply have to revert γ and use the previous condition for γ_c. In particular, we obtain that (8) still holds, where now $i^{\triangleleft}$, $i^{\triangleright}$, $s_{i^{\triangleleft}}$ and $s_{i^{\triangleright}}$ are evaluated at any point in (γ_c, γ') for some γ' close to γ_c. In addition, if the ith point transitions from non-free to free, then

$$\tilde{\alpha}_{i^{\triangleleft}} s_i s_{i^{\triangleleft}} f_i + \tilde{\alpha}_{i^{\triangleright}} s_i s_{i^{\triangleright}} \bar{f}_i < \tilde{\alpha}_i. \tag{10}$$

From the above observations, we can prove bounds on T.

• 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, as far as we know, 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 non-free, then, by (9), we must have that

$$1 < s_i s_{i^{\triangleleft}} f_i + s_i s_{i^{\triangleright}} \bar{f}_i. \tag{11}$$

This implies that

$$1 < |s_i s_{i^{\triangleleft}} f_i + s_i s_{i^{\triangleright}} \bar{f}_i| \leq f_i + \bar{f}_i = 1, \tag{12}$$

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 (see our second observation), and since F can only change by the addition of $n-1$ variables at most, we have that $T = \mathcal{O}(n)$. □

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

Proof. It is enough to prove that w^* has at most $\mathcal{O}(n)$ different linear segments. A change in linear segment occurs in $w^*(\gamma)$ when, as γ changes, some i changes from B to F. Any value of γ at which i transitions from B to F (or vice-versa) satisfies (8) for some value of $i^{\triangleright}$, $i^{\triangleleft}$, $s_{i^{\triangleleft}}$ and $s_{i^{\triangleright}}$. There are $\mathcal{O}(n^2)$ choices for these values and hence $\mathcal{O}(n^2)$ possible values of γ at which i might transition. Since there are n components in w^*, there are $\mathcal{O}(n^3)$ values of γ at which B might change and hence $\mathcal{O}(n^3)$ different linear segments. □

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

$$\tilde{\alpha}_i = (i-1)^2, \forall i \in [n], \tilde{\alpha}_{n+1} = 0, \tag{13}$$

$$\tilde{y}_i = (-1)^i q_i \forall i \in [n], \tilde{y}_{n+1} = 0, \tag{14}$$

$$q_1 = 1, q_2 = 2, q_{i+2} = 2q_{i+1} - q_i + 2g_{i+2} + 1, \forall i \in [n-2], \tag{15}$$

$$g_3 = 1/3, g_{i+3} = 2g_{i+2} + 1. \tag{16}$$

Remark 5. *This automatically implies that there are examples for which variables "fuse" $\Omega(n^2)$ times and "un-fuse" $\Omega(n^2)$ times.*

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 its 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 $\tilde{\alpha}_2, \ldots, \tilde{\alpha}_n \neq 0$ 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:

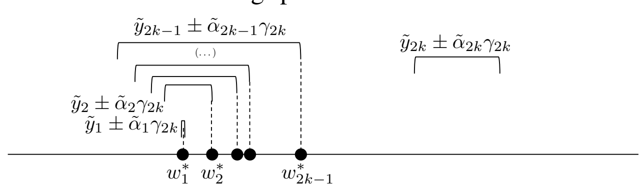
1) for $\gamma = \gamma_{2k}$, $3 \leq 2k \leq n$, any point w_i^* with $i \in \{2, \ldots, 2k-1\}$ is touching its right boundary;
2) for $\gamma = \gamma_{2k+1}$, $3 \leq 2k+1 \leq n$, any point w_i^* with $i \in \{2, \ldots, 2k\}$ is touching its left boundary;

Both scenarios imply that, for $\gamma \in [\gamma_r, \gamma_{r+1})$, $r \in \{3, \ldots, n-1\}$, any 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 $\tilde{\alpha}_2, \ldots \tilde{\alpha}_n \neq 0$, 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

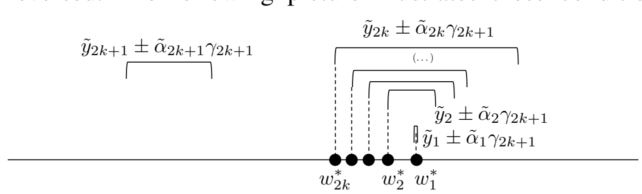
$$1 + \cdots + n - 1 = n(n-1)/2 \tag{17}$$

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 $[\tilde{y}_1 - \tilde{\alpha}_1\gamma, \tilde{y}_1 + \tilde{\alpha}_1\gamma]$, $\ldots$, $[\tilde{y}_{r-1} - \tilde{\alpha}_{r-1}\gamma, \tilde{y}_{r-1} + \tilde{\alpha}_{r-1}\gamma]$ 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 $r \in \{3, \ldots, n\}$,

$$|\tilde{y}_{i+1} - \tilde{y}_i| < (\tilde{\alpha}_{i+1} - \tilde{\alpha}_i)\gamma_r, \forall i \in [r-2], \tag{18}$$

$$|\tilde{y}_{i+2} - 2\tilde{y}_{i+1} + \tilde{y}_i| < (\tilde{\alpha}_{i+2} - 2\tilde{\alpha}_{i+1} + \tilde{\alpha}_i)\gamma_r, \forall i \in [r-3], \tag{19}$$

$$(-1)^r(\tilde{y}_r - 2\tilde{y}_{r-1} + \tilde{y}_{r-2}) > (\tilde{\alpha}_r + 2\tilde{\alpha}_{r-1} - \tilde{\alpha}_{r-2})\gamma_r. \tag{20}$$

Condition (18) directly implies that the first $r-1$ intervals are nested. Condition (19) 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. In terms of our physical interpretation using springs, condition (19) implies that the length of the spring connecting the $(i+2)$th to the $(i+1)$th point is larger than the length of the spring connecting the $(i+1)$th to the ith point. Hence, the force that the ith point feels from its left and right springs pushes the point to the right (left) so that it touches its right (left) limit. Condition (20) 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 has a physical interpretation similar to that of condition (19). 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.

$$\tilde{\alpha}_{i+2} - 2\tilde{\alpha}_{i+1} + \tilde{\alpha}_i > 0, \forall i \in [n-2], \tag{21}$$
$$\tilde{\alpha}_{i+1} - \tilde{\alpha}_i > 0, \forall i \in [n-1], \tag{22}$$
$$\gamma_3 > \frac{q_2 - q_1}{\tilde{\alpha}_2 - \tilde{\alpha}_1}, \tag{23}$$
$$\gamma_{i+3} > \max\left\{\gamma_{i+2}, \frac{q_{i+2} - 2q_{i+1} + q_i}{\tilde{\alpha}_{i+2} - 2\tilde{\alpha}_{i+1} + \tilde{\alpha}_i}, \frac{q_{i+2} - q_{i+1}}{\tilde{\alpha}_{i+2} - \tilde{\alpha}_{i+1}}\right\}, \quad \forall i \in [n-3], \tag{24}$$
$$q_2 > q_1, \tag{25}$$
$$q_{i+2} > 2q_{i+1} - q_i + (\tilde{\alpha}_{i+2} + 2\tilde{\alpha}_{i+1} - \tilde{\alpha}_i)\gamma_{i+2}, \forall i \in [n-2], \tag{26}$$
$$\tilde{y}_i = (-1)^i q_i, \forall i \in [n]. \tag{27}$$

Conditions (22), (23) and (25) imply that $\gamma_3 > 0$. Condition (24) further implies that $0 < \gamma_3 < \cdots < \gamma_n$. Condition (26) and (27) imply condition (20). Since $\tilde{\alpha} > 0$, condition (21) 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 (25) together with (26) 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 (21), (22), (23) and (24) imply that (18) holds for all $i \in [n-1]$, if we replace γ_r by γ_{i+2}. These also imply that (19) 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 both (18) and (19) hold true exactly as specified.

We finally argue by contradiction why conditions (21)-(27) also imply that no two consecutive components of w^* stop touching, or start touching, its 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 (17).

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 (17). 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 (21)-(27) imply that $|\tilde{y}_{i+1} - 2\tilde{y}_i + \tilde{y}_{i-1}| < (\tilde{\alpha}_{i+1} - 2\tilde{\alpha}_i + \tilde{\alpha}_{i-1})\gamma_r \leq (\tilde{\alpha}_{i+1} - 2\tilde{\alpha}_i + \tilde{\alpha}_{i-1})\gamma_c$, 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 (13)-(16). As Theorem 4 predicts, the number of critical points grows quadratically with 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 αs, and each y_i was sampled from a $\mathcal{N}(0, \sqrt{10})$, independently across ys. Although Theorem 4 tells us that we can observe $\Omega(n^2)$ “fuse” and “un-fuse” events, in our random instances for W1FL, “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 a very simple algorithm to compute the critical points based on Theorem 1.

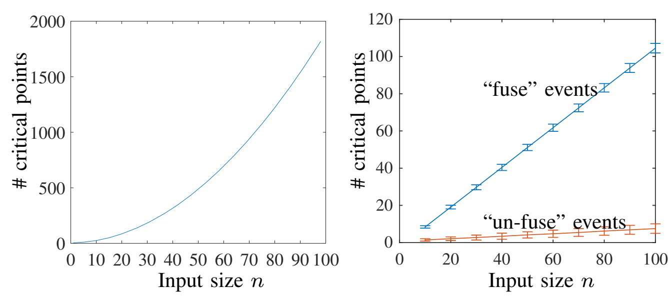


Fig. 3. (Left) Number of critical points as function of n for the example of Theorem 4; (Right) Number of “fuse” and “un-fuse” events in random instances of W1FL.

IV. Future work

The weighted fused lasso on an input of size n is substantially different from the equal-weights fused lasso in that 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 $\mathcal{O}(n^3)$ events. We have also produced a very simple proof of why in the equal-weights fused lasso there are $\mathcal{O}(n)$ events at most.

Future work should include closing this gap. We conjecture that it is our upper bound that needs improvement. Furthermore, since random instances have a number of events that grows linearly with n, it would be good to find conditions for the weights α and input y under which the $\mathcal{O}(n)$ growth holds and to compute how likely it is, under different stochastic models, for these conditions to be satisfied.

Acknowledgment

We would like to thank Piotr Suwara for his important help in some of the proofs. This work was partially supported by the grants NSF-IIS-1741129 and NIH-1U01AI124302.

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

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

be its proximal operator. Let

$$\hat{f}(x) = \sup_z \langle z, x\rangle - f(z) \tag{29}$$

be the Fenchel dual of f and let

$$\hat{F}(y) = \arg\min_x \hat{f}(x) + \frac{1}{2}\|x-y\|_2^2 \tag{30}$$

be its proximal operator. Moreau's identity is

$$F(y) = y - F^*(y). \tag{31}$$

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 (31).

We start by making a change of variables in

$$f^*(x) = \sup_z \sum_{t=1}^{n} z_t x_t - \sum_{t=1}^{n-1} \alpha_t |z_{t+1} - z_t|. \tag{32}$$

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 (32) as

$$f^*(x) = \sup_h \sum_{t=0}^{n-1} h_t u_t - \sum_{t=0}^{n-1} \gamma\alpha_t |h_t|, \tag{33}$$

where we have extended α such that $\alpha_0 = 0$. Problem (33) breaks down into n independent one-dimensional problems of the form

$$\sup_{h_t} h_t u_t - \gamma\alpha_t |h_t|, \tag{34}$$

that have solution 0 if $u_t \in [-\gamma\alpha_t, \gamma\alpha_t]$, and ∞ otherwise. Therefore, $f^*(x) = \infty$ 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 $f^*(x) = 0$ otherwise.

We can now write

$$F(y) = y - F^*(y) = y - \arg\min_x \sum_{t=1}^{n} (x_t - y_t)^2 \tag{35}$$

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

Finally, we make the following change of variable in (35): $w_t = \sum_{i=t}^{n} (x_i - y_i)$ for all $t \in [n+1]$, where we define $w_{n+1} = 0$. This implies that $x_i - y_i = y_i + w_{i+1} - w_i$ for $i \in [n]$, and hence that

$$x_i^* = F(y)_i = y_i - F^*(y)_i = y_i - (y_i + w_{i+1}^* - w_i^*) \tag{36}$$

$$= w_{i+1}^* - w_i^* \tag{37}$$

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

$$w^* = \arg\min_w \sum_{t=1}^{n} (w_{t+1} - w_t)^2 \tag{38}$$

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

□

VI. Simple algorithm to compute the critical points

Before we introduce the algorithm, we need a fifth observation, in addition to the four observations already made in the main text. For this observation to be valid, we assume that the αs are not in some set of measure zero on the space of all possible αs. The following technical results, whose proofs are standard, e.g. [28], [29], will be used to extend these observations to any set of αs. 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 $w^*(\gamma, \alpha)_{t+1} = \tilde{y}_1 + \sum_{i=1}^{t} x^*(\gamma, \alpha)_i$, this also proves that $w^*(\gamma, \alpha)$ is continuous at every point of its domain.*

The fifth observation is that we can assume, without loss of generality, that at most one component of w^* transitions from free to non-free, or vice versa, at each γ. This follows from the fact that, for two indices i and j (or more) to satisfy (8) 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 αs). Using Lemma 6 and Lemma 7, we can then extend our arguments made outside ζ to this set as well.

Our five 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 springs relax, 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 (7) to compute how each point moves and hence compute the next B.

Let us be more precise [2].

1) Start with $\gamma_1 = 0$, $F = \{\}$ and iteration number $r = 1$. All points are non-free.
2) At iteration r, use (8) to find γs at which an $i \notin F$ *might* enter F or at which an $i \in F$ *might* leave F. Store these values as $\gamma_c^{(1)}$,$\gamma_c^{(2)}$, ... and $\gamma_c'^{(1)}$, $\gamma_c'^{(2)}$, ... 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. These values are not all critical points on $w^*(\gamma)$, but all critical points satisfy (8),

[2] The following algorithm only describes how to compute $\gamma_1, \ldots, \gamma_T$, the critical points. It can be modified to compute $w^*(\gamma_1), \ldots, w^*(\gamma_T)$ with only a multiplying factor slow down in the run time.

and, in particular, the smallest value produced corresponds to the next critical point.

3) Append or remove from F the index

$$i_c = \min\{\min_{i:\gamma_c^{(i)}>\gamma_r} \gamma_c^{(i)}, \min_{i:\gamma_c^{\prime(i)}>\gamma_r} \gamma_c^{\prime(i)}\}. \tag{39}$$

Set $\gamma_{r+1} = \gamma_c^{(i_c)}$ or $\gamma_{r+1} = \gamma_c^{\prime(i_c)}$. Set $\gamma_c^{(i)} = \infty$ or $\gamma_c^{\prime(i)} = \infty$, the choice depending on whether there is a point leaving, or entering F.

4) Update $\gamma_c^{(j)}$ or $\gamma_c^{\prime(j)}$ for all $j \in \{i^{\triangleleft}, \ldots, i-1, i+1, \ldots, i^{\triangleright}\}$. All the other previously computed $\gamma_c^{(j)}$ and $\gamma_c^{\prime(j)}$ 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 an i, either in B or in F, this allows us to determine in $\mathcal{O}(1)$ steps the indices $i^{\triangleleft}$ and $i^{\triangleright}$. 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 $\{\gamma_c^{\prime(j)}\}$, we pay a computational cost of $\mathcal{O}(\log n)$ each time we update $\gamma_c^{(j)}$ or $\gamma_c^{\prime(j)}$ for some j. Therefore, the complexity of this algorithm is

$$\mathcal{O}\left(\left(\sum_{r=1}^{T} k_r\right) \log n\right) \tag{40}$$

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 (8) 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 $\mathcal{P}(\alpha) = \{z \in \mathbb{R}^n : |\tilde{y}_i - \sum_{r=1}^{i} z_r| \leq \gamma\tilde{\alpha}_{i-1}$ 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.

□