A PROOF OF THEOREM 3.1

Our proof of Theorem 3.1 is structured as follows. First, we shows that any $f \in \text{RidgelessReLU}(\boldsymbol{D})$ satisfies properties (1) and (2). This constitutes the majority of the argument and requires several preparatory results, starting with Proposition A.1 and its Corollary A.3. With these in hand, we derive in Propositions A.7, A.9, and A.11 constraints on the local behavior of f on small intervals of the form (x_i, x_{i+1}) or (x_{i-1}, x_{i+1}). Taken together these Propositions, and several other results, imply properties (1) and (2). The details for this step are around Lemma A.12. Finally, establish in Proposition A.13 that any f which satisfies properties (1) and (2) belongs to $\text{RidgelessReLU}(\boldsymbol{D})$. To start, we introduce some notation. For each $f \in \text{PL(D)}$ and every $x \in \mathbb{R}$, let us write

$$s_{\text{in}}(x) = s_{\text{in}}(f, x) := \lim_{\epsilon \to 0^+} Df(x - \epsilon), \qquad s_{\text{out}}(x) = s_{\text{out}}(f, x) := \lim_{\epsilon \to 0^+} Df(x + \epsilon)$$

for the incoming and outgoing slopes of f at x. For any $f \in \text{PL}$ the second derivative $D^2 f$ is an atomic measure and we have

$$D^2 f = \sum_{j=1}^{k} c_j \delta_{\xi_j}, \qquad c_j = s_{\text{out}}(f, \xi_j) - s_{\text{in}}(f, \xi_j)$$

where ξ_j are the points of discontinuity for the derivative Df. We will usually supress f from the notation. Thus, Df, and in particular Dz for any one layer ReLU network z, has a well-defined total variation

$$||Df||_{TV} := \sum_{j=1}^{k} |c_j| \,.$$

Much of the remainder of our proof results on the following fundamental observation.

Proposition A.1. *Fix $f \in \text{RidgelessReLU}(\boldsymbol{D})$. For every $i = 1, \ldots, m-1$ and Df is monotone on (x_i, x_{i+1}) in the sense that the functions $s_{\text{in}}(f, x)$ and $s_{\text{out}}(f, x)$ are both either non-increasing or non-decreasing for $x \in (x_i, x_{i+1})$.*

Proof. We proceed by contradiction. That is, let us suppose that $f \in \text{RidgelessReLU}(\boldsymbol{D})$ and that for some i there exist

$$x_i \leq \xi_1 < \xi_2 < \xi_3 < \xi_4 \leq x_{i+1}$$

such that f is given by distinct affine functions with slopes σ_j when restricted to any of (ξ_j, ξ_{j+1}) for $j = 1, 2, 3$ but that the sequence $\sigma_1, \sigma_2, \sigma_3$ is not monotone. Without loss of generality we assume

$$\sigma_1, \sigma_3 < \sigma_2. \tag{9}$$

In particular, for all δ sufficiently small, we have

$$\text{Total Variation of } Df \text{ on } (\xi_1 - \delta, \xi_4 + \delta) = 2\sigma_2 - \sigma_1 - \sigma_3 + |\sigma_1 - \sigma_{\text{in}}| + |\sigma_3 - \sigma_{\text{out}}| \,, \tag{10}$$

where

$$\sigma_{\text{in}} := s_{\text{in}}(f, \xi_1) = \lim_{\epsilon \to 0^+} Df(\xi_1 - \epsilon)$$

and

$$\sigma_{\text{out}} := s_{\text{out}}(f, \xi_4) = \lim_{\epsilon \to 0^+} Df(\xi_4 + \epsilon).$$

Define

$$\sigma_* := \frac{f(\xi_4) - f(\xi_1)}{\xi_1 - \xi_4} = \frac{\sigma_1(\xi_2 - \xi_1) + \sigma_2(\xi_3 - \xi_2) + \sigma_3(\xi_4 - \xi_3)}{\xi_4 - \xi_1}.$$

Note that the constraint (9) and the fact that σ_* is a convex combination of σ_j guarantees that

$$\min\{\sigma_1, \sigma_3\} < \sigma_* < \sigma_2. \tag{11}$$

See Figure 4 for the three possible cases. Consider $g \in \text{PL(D)}$ defined as follows:

$$g(x) = \begin{cases} f(x), & x \in (\xi_1, \xi_4)^c \\ \sigma_*(x - \xi_1) + f(\xi_1), & x \in (\xi_1, \xi_4) \end{cases}.$$

The function g represents a "straightening of f" between ξ_1 and ξ_4, and we will now show that the total variation of Dg on $(\xi_1 - \delta, \xi_4 + \delta)$ is strictly smaller than that of Df on the same interval. Since

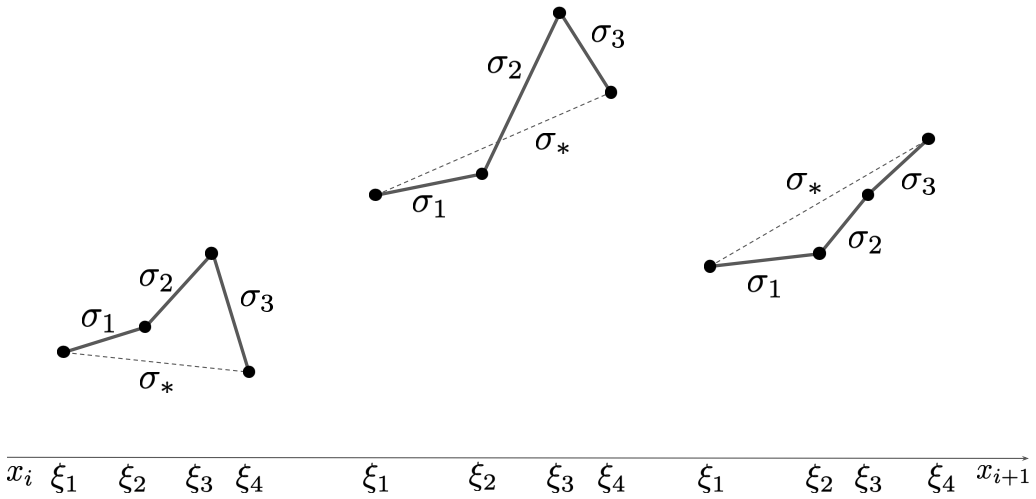

Figure 4: The three possible relative configurations for σ_j, σ_* are shown. On the left, $\sigma_3 < \sigma_* \leq \sigma_1 < \sigma_2$. In the center $\sigma_1, \sigma_3 < \sigma_* < \sigma_2$. On the right, $\sigma_1 < \sigma_* < \sigma_3 < \sigma_2$.

the total variations of Df and Dg agree on $(\xi_1, \xi_4)^c$ this will contradict the minimality of $||Df||_{TV}$ over $\mathrm{PL(D)}$. Indeed, considering all possible cases for the relative sizes of $\sigma_{\text{in}}, \sigma_{\text{out}}$ and σ_* we find for all δ sufficiently small

$$\text{Total Variation of } Dg \text{ on } (\xi_1 - \delta, \xi_4 + \delta) = \max\left\{|2\sigma_* - \sigma_{\text{in}} - \sigma_{\text{out}}|, |\sigma_{\text{in}} - \sigma_{\text{out}}|\right\}.$$

Combining this with the expression (10) for the total variation of Df and the following elementary Lemma completes the proof.

Lemma A.2. *For any $\sigma_1, \sigma_2, \sigma_3, \sigma_*$ satisfying (11) we have*

$$2\sigma_2 - \sigma_1 - \sigma_3 + |\sigma_1 - \sigma_{\text{in}}| + |\sigma_3 - \sigma_{\text{out}}| > \max\left\{|2\sigma_* - \sigma_{\text{in}} - \sigma_{\text{out}}|, |\sigma_{\text{in}} - \sigma_{\text{out}}|\right\}.$$

Proof. We consider all four cases for the maximum on the right hand side. We have

$$\begin{aligned}
&2\sigma_2 - \sigma_1 - \sigma_3 + |\sigma_1 - \sigma_{\text{in}}| + |\sigma_3 - \sigma_{\text{out}}| - (2\sigma_* - \sigma_{\text{in}} - \sigma_{\text{out}}) \\
&\quad = 2(\sigma_2 - \sigma_*) + |\sigma_1 - \sigma_{\text{in}}| - (\sigma_1 - \sigma_{\text{in}}) + |\sigma_3 - \sigma_{\text{out}}| - (\sigma_3 - \sigma_{\text{out}}) \\
&\quad > 0,
\end{aligned}$$

as desired. Similarly,

$$\begin{aligned}
&2\sigma_2 - \sigma_1 - \sigma_3 + |\sigma_1 - \sigma_{\text{in}}| + |\sigma_3 - \sigma_{\text{out}}| - (\sigma_{\text{in}} + \sigma_{\text{out}} - 2\sigma_*) \\
&\quad = 2(\sigma_2 + \sigma_* - \sigma_1 - \sigma_3) + |\sigma_1 - \sigma_{\text{in}}| - (\sigma_{\text{in}} - \sigma_1) + |\sigma_3 - \sigma_{\text{out}}| - (\sigma_{\text{out}} - \sigma_3) \\
&\quad > 0,
\end{aligned}$$

as desired. Further,

$$\begin{aligned}
&2\sigma_2 - \sigma_1 - \sigma_3 + |\sigma_1 - \sigma_{\text{in}}| + |\sigma_3 - \sigma_{\text{out}}| - (\sigma_{\text{in}} - \sigma_{\text{out}}) \\
&\quad = 2(\sigma_2 - \sigma_1) + |\sigma_1 - \sigma_{\text{in}}| - (\sigma_{\text{in}} - \sigma_1) + |\sigma_3 - \sigma_{\text{out}}| - (\sigma_3 - \sigma_{\text{out}}) \\
&\quad > 0,
\end{aligned}$$

as desired. Finally,

$$\begin{aligned}
&2\sigma_2 - \sigma_1 - \sigma_3 + |\sigma_1 - \sigma_{\text{in}}| + |\sigma_3 - \sigma_{\text{out}}| - (\sigma_{\text{out}} - \sigma_{\text{in}}) \\
&\quad = 2(\sigma_2 - \sigma_3) + |\sigma_1 - \sigma_{\text{in}}| - (\sigma_1 - \sigma_{\text{in}}) + |\sigma_3 - \sigma_{\text{out}}| - (\sigma_{\text{out}} - \sigma_3) \\
&\quad > 0,
\end{aligned}$$

completing the proof. □

□

Proposition A.1 shows that any $f \in \mathrm{RidgelessReLU}(\boldsymbol{D})$ is either convex or concave on any interval of the form (x_i, x_{i+1}). This gives several useful consequences, for example the following

Corollary A.3 (of Proposition A.1)**.** *Fix $f \in \mathrm{RidgelessReLU}(\boldsymbol{D})$. For each $i = 1, \ldots, m-1$,*

$$\operatorname{sgn}\left(s_{\mathrm{in}}(x_{i+1}) - s_i\right) + \operatorname{sgn}\left(s_{\mathrm{out}}(x_i) - s_i\right) = 0.$$

Proof. Suppose first $\operatorname{sgn}(s_{\mathrm{out}}(x_i) - s_i) = 0$. That is, $s_{\mathrm{out}}(x_i) = s_i$. By Proposition A.1 we have $s_{\mathrm{out}}(x)$ is monotone for $x \in (x_i, x_{i+1})$. Thus, if there exists $\xi \in (x_i, x_{i+1})$ so that $Df(\xi) > s_i$, then $f(x_{i+1}) > (x_{i+1} - x_i)s_i + f(y_i) = y_{i+1}$, contradicting the assumption that $f \in \mathrm{PL(D)}$. A similar contradiction occurs if there exists $\xi \in (x_i, x_{i+1})$ so that $Df(\xi) < s_i$. Hence, we conclude that $s_{\mathrm{in}}(x_{i+1}) = s_i$, as desired. Next, suppose $s_{\mathrm{out}}(x_i) > s_i$. In particular, there exists $\xi_+ \in (x_i, x_{i+1})$ such that

$$s_{\mathrm{in}}(f, \xi_+) > s_i.$$

Since f satisfies $f(x_i) = y_i$ and $f(x_{i+1}) = y_{i+1}$ there must exist $\xi_- \in (x_i, x_{i+1})$ such that

$$s_{\mathrm{in}}(f, \xi_-) < s_i.$$

By Proposition A.1, $s_{\mathrm{in}}(f, \xi)$ is monotone for $\xi \in (x_i, x_{i+1})$. We see by comparing $s_{\mathrm{in}}(f, \xi_\pm)$ that it is in fact non-increasing. Since $x_{i+1} - \delta > \xi_-$ for δ sufficiently small, we conclude that $s_{\mathrm{in}}(x_{i+1}) < s_i$, as desired. The case $s_{\mathrm{out}}(x_i) < s_i$ is analogous, completing the proof. □

For the remainder of the proof we fix $f \in \mathrm{RidgelessReLU}(\boldsymbol{D})$ and show that it must satisfy properties (1) and (2). To prove this, we use Proposition A.1 and Corollary A.3 to derive Propositions A.4, A.5, A.7, and A.9 that together determine the structure of f. Specifically, Propositions A.4, A.5 and a combination of Propositions A.7 and A.9 show that f satisfies property (1). Then, a different application of Propositions A.7 and A.9, together with the fact that f satisfies property (1), will imply that f satisfies property (2) as well.

Proposition A.4 (f agrees with $f_{\boldsymbol{D}}$ on colinear neighbors)**.** *Fix $i = 2, \ldots, m-1$. Suppose $\epsilon_i = 0$. Then*

$$s_{\mathrm{out}}(x_{i-1}) = s_{\mathrm{in}}(x_i) = s_{\mathrm{out}}(x_i) = s_{\mathrm{in}}(x_{i+1}) = s_{i-1} = s_i.$$

Hence, $f(x) = f_{\boldsymbol{D}}(x)$ for all $x \in (x_{i-1}, x_{i+1})$.

Proof. By definition, since $\epsilon_i = 0$, we have $s_i = s_{i-1}$. Suppose for the sake of contradiction that

$$\text{at least one of } s_{\mathrm{out}}(x_{i-1}),\ s_{\mathrm{in}}(x_i),\ s_{\mathrm{out}}(x_i),\ s_{\mathrm{in}}(x_{i+1}) \text{ does not equal } s_i.$$

By Corollary A.3, this means that either one or both least one of the pairs $(s_{\mathrm{out}}(x_{i-1}),\ s_{\mathrm{in}}(x_i))$ or $(s_{\mathrm{out}}(x_i),\ s_{\mathrm{in}}(x_{i+1}))$ are both not equal to s_i. We will suppose without loss of generality that

$$\min\left\{s_{\mathrm{out}}(x_{i-1}), s_{\mathrm{in}}(x_i)\right\} < s_i < \max\left\{s_{\mathrm{out}}(x_{i-1}), s_{\mathrm{in}}(x_i)\right\}. \tag{12}$$

Note also that by Corollary A.3 and the fact that $f(x_i) = y_i$ and $f(x_{i+1}) = y_{i+1}$ we also have

$$\min\left\{s_{\mathrm{out}}(x_i), s_{\mathrm{in}}(x_{i+1})\right\} \le s_i \le \max\left\{s_{\mathrm{out}}(x_i), s_{\mathrm{in}}(x_{i+1})\right\}. \tag{13}$$

By definition, if $\epsilon_i = 0$, then $s_{i-1} = s_i$. By Proposition A.1, the total variation of Df on $(x_{i-1} - \delta, x_{i+1} + \delta)$ equals, for all δ sufficiently small,

$$\begin{aligned}&|s_{\mathrm{out}}(x_{i+1}) - s_{\mathrm{in}}(x_{i+1})| + |s_{\mathrm{in}}(x_{i+1}) - s_{\mathrm{out}}(x_i)| + |s_{\mathrm{out}}(x_i) - s_{\mathrm{in}}(x_i)| \\ &\quad + |s_{\mathrm{in}}(x_i) - s_{\mathrm{out}}(x_{i-1})| + |s_{\mathrm{out}}(x_{i-1}) - s_{\mathrm{in}}(x_{i-1})|,\end{aligned}$$

which is bounded below by

$$|s_{\mathrm{out}}(x_{i+1}) - s_{\mathrm{in}}(x_{i+1})| + |s_{\mathrm{in}}(x_{i+1}) - s_{\mathrm{out}}(x_i)| + |s_{\mathrm{in}}(x_i) - s_{\mathrm{out}}(x_{i-1})| + |s_{\mathrm{out}}(x_{i-1}) - s_{\mathrm{in}}(x_{i-1})|.$$

Define $g \in \mathrm{PL(D)}$ to coincide with f on $(x_{i-1}, x_{i+1})^c$ and to coincide with $f_{\boldsymbol{D}}$ on (x_{i-1}, x_{i+1}). The total variation of Dg on $(x_{i-1} - \delta, x_{i+1} + \delta)$ equals, for all δ sufficiently small,

$$|s_{\mathrm{out}}(x_{i+1}) - s_i| + |s_{\mathrm{in}}(x_{i-1}) - s_i|.$$

Using that

$$|s_{\text{out}}(x_{i+1}) - s_i| \leq |s_{\text{out}}(x_{i+1}) - s_{\text{in}}(x_{i+1})| + |s_{\text{in}}(x_{i+1}) - s_i|$$

and

$$|s_{\text{in}}(x_{i-1}) - s_i| \leq |s_{\text{in}}(x_{i-1}) - s_{\text{out}}(x_{i-1})| + |s_{\text{out}}(x_{i-1}) - s_i|\,,$$

we find that the difference between the total variation of Df and Dg on $(x_{i-1} - \delta, x_{i+1} + \delta)$ is bounded below by

$$|s_{\text{in}}(x_{i+1}) - s_{\text{out}}(x_i)| - |s_{\text{in}}(x_{i+1}) - s_i| + |s_{\text{in}}(x_i) - s_{\text{out}}(x_{i-1})| - |s_i - s_{\text{out}}(x_{i-1})|\,.$$

Note that if $a, c \in \mathbb{R}$ and $\min\{a, c\} \leq b \leq \max\{a, c\}$, then we have

$$|c - a| - |a - b| = |b - c|\,.$$

Hence, using our assumptions (12) and (13), we conclude that

$$|s_{\text{in}}(x_i) - s_{\text{out}}(x_{i-1})| - |s_i - s_{\text{out}}(x_{i-1})| = |s_{\text{in}}(x_i) - s_i| > 0$$

and that

$$|s_{\text{in}}(x_{i+1}) - s_{\text{out}}(x_i)| - |s_{\text{in}}(x_{i+1}) - s_i| = |s_{\text{in}}(x_{i+1}) - s_i| \geq 0.$$

The difference between the total variation of Df and Dg on $(x_{i-1} - \delta, x_{i+1} + \delta)$ is thus strictly positive for all δ sufficiently small. Since f, g agree on $(x_{i-1}, x_{i+1})^c$, we find that $||Dg||_{TV} < ||Df||_{TV}$, contradicting the minimality of $||Df||_{TV}$ over PL(D). □

Our next result, Proposition A.5, ensures that f and $f_{\boldsymbol{D}}$ agree near infinity.

Proposition A.5. *Suppose $f \in \text{RidgelessReLU}(\boldsymbol{D})$. Then for $x < x_2$ and $x > x_{m-1}$ we have that $f(x) = f_{\boldsymbol{D}}(x)$.*

Proof. We focus on the analysis of f on $(-\infty, x_2)$ since the conclusion on (x_{m-1}, ∞) follows by symmetry. To start note that $Df(x) = s_{\text{out}}(x_1)$ for all $x < x_1$. Indeed, if this were not the case, we could define $g \in \text{PL(D)}$ to coincide with f on (x_1, ∞) but to have slope $s_{\text{out}}(x_1)$ on $(-\infty, x_1)$. This g belongs to PL(D) and satisfies $||Dg||_{TV} < ||Df||_{TV}$ since the total variation of its derivative on $(-\infty, x_1 + \epsilon)^c$ equals that of Df but the total variation of Dg on $(-\infty, x_1 + \epsilon)$ vanishes while that of f is non-zero.

Thus, we see that $s_{\text{in}}(x_1) = s_{\text{out}}(x_1)$. Let us now prove that $f(x) = f_{\boldsymbol{D}}(x)$ for $x \in (x_1, x_2)$. This will imply $s_{\text{out}}(x_1) = s_1$ and will complete the proof. Suppose for the sake of contradiction that $s_{\text{in}}(x_2) \neq s_1$. Then we have from Corollary A.3 that

$$\min\{s_{\text{out}}(x_1), s_{\text{in}}(x_2)\} < s_1 < \max\{s_{\text{out}}(x_1), s_{\text{in}}(x_2)\}\,.$$

Define $g \in \text{PL(D)}$ to coincide with f on (x_2, ∞) and with $f_{\boldsymbol{D}}$ on $(-\infty, x_2)$. The total variation of Dg on $(-\infty, x_2 + \delta)$ for all δ sufficiently small is

$$|s_{\text{out}}(x_2) - s_1|\,,$$

whereas the total variation of Df on the same interval is

$$|s_{\text{out}}(x_1) - s_{\text{in}}(x_2)| + |s_{\text{in}}(x_2) - s_{\text{out}}(x_2)|\,.$$

Since by construction Df and Dg agree on (x_2, ∞), the following claim shows that $||Df||_{TV} > ||Dg||_{TV}$, contradicting the minimality of $||Df||_{TV}$ over PL(D):

Claim A.6. *Suppose $a, b, c \in \mathbb{R}$ satisfy*

$$\min\{a, b\} < c < \max\{a, b\}\,.$$

Then for any $d \in \mathbb{R}$ we have

$$|d - c| < |a - b| + |b - d|$$

Proof. Suppose first $a < c < b$. Then

$$\begin{aligned} |a - b| + |b - d| - |d - c| &= b - a + |b - d| - |d - c| \\ &= c - a + |b - d| - (d - b) - |d - c| + (d - c) \\ &> 0, \end{aligned}$$

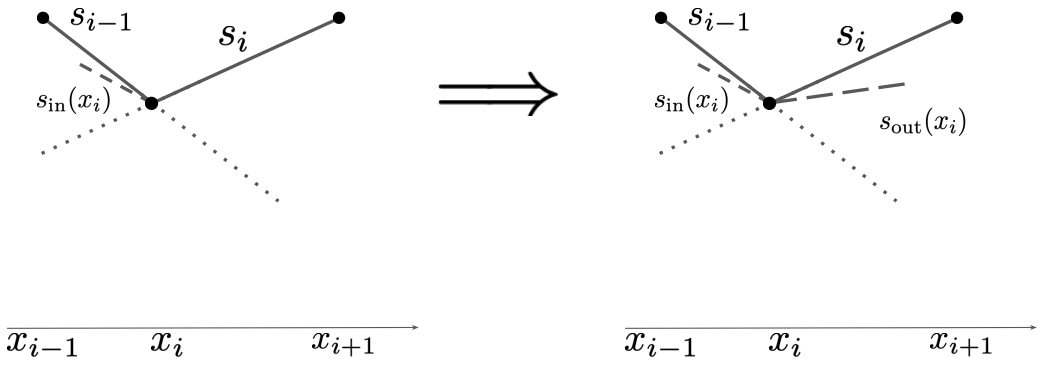

Figure 5: The conclusion of Proposition A.7 when $\epsilon_i = 1$.

as desired. Similarly, suppose $b < c < a$ then

$$|a-b| + |b-d| - |d-c| = a - b + |b-d| - |d-c|\,. \tag{14}$$

If $d \geq c$ then $d > b$ and the right hand side of (14) becomes

$$a - b + |b-d| - (d-c) = a - b + d - b - d + c = c - b + a - b > 0.$$

Finally, if $d \leq c$ then the right hand side of (14) becomes

$$a - b + |d-b| - (c-d) = a - c + |d-b| - (d-b) > 0.$$

This completes the proof. □

□

Proposition A.5 allows us to know the "initial" and "final" conditions $s_{\text{in}}(x_2)$ and $s_{\text{out}}(x_{m-1})$ for the slopes of f. In contrast, Proposition A.7 below allows us to take information about the incoming slope $s_{\text{in}}(x_i)$ of f at x_i and use the local curvature information ϵ_i at x_i to constrain the outgoing slope $s_{\text{out}}(x_i)$. See Figure 5.

Proposition A.7 (How slope of f changes at x_i)**.** *Suppose* $\epsilon_i = 1$. *Then*

$$s_{i-1} \leq s_{\text{in}}(x_i) \leq s_i \quad \Longrightarrow \quad s_{i-1} \leq s_{\text{in}}(x_i) \leq s_{\text{out}}(x_i) \leq s_i \tag{15}$$

Similarly, suppose $\epsilon_i = -1$. *Then*

$$s_{i-1} \geq s_{\text{in}}(x_i) \geq s_i \quad \Longrightarrow \quad s_{i-1} \geq s_{\text{in}}(x_i) \geq s_{\text{out}}(x_i) \geq s_i \tag{16}$$

Proof. The proof of (16) is identical to that of (15), and we therefore focus on proving the latter. That is, we fix $i = 2, \ldots, m-1$ and assume $\epsilon_i = 1$ and suppose that $s_{i-1} \leq s_{\text{in}}(x_i) \leq s_i$. For the sake of contradiction assume also that $s_{\text{out}}(x_i) > s_i$. By Corollary A.3 we have $s_{\text{in}}(x_{i+1}) < s_i$ and therefore the total variation of Df on $(x_i - \epsilon, x_{i+1} + \epsilon)$ is

$$|s_{\text{out}}(x_{i+1}) - s_{\text{in}}(x_{i+1})| + 2s_{\text{out}}(x_i) - s_{\text{in}}(x_{i+1}) - s_{\text{in}}(x_i).$$

Consider $g \in \text{PL(D)}$ defined to be equal to f on $(x_i, x_{i+1})^c$ and to $f_{\boldsymbol{D}}$ on (x_i, x_{i+1}). The total variation of Dg on $(x_i - \delta, x_{i+1} + \delta)$ for all δ sufficiently small is

$$|s_{\text{out}}(x_{i+1}) - s_i| + s_i - s_{\text{in}}(x_i).$$

The following claim shows that the total variation of Dg on $(x_i - \delta, x_{i+1} + \delta)$ for all δ sufficiently small is strictly smaller than that of Df. Implies that $||Dg||_{TV} < ||Df||_{TV}$, which is a contradiction.

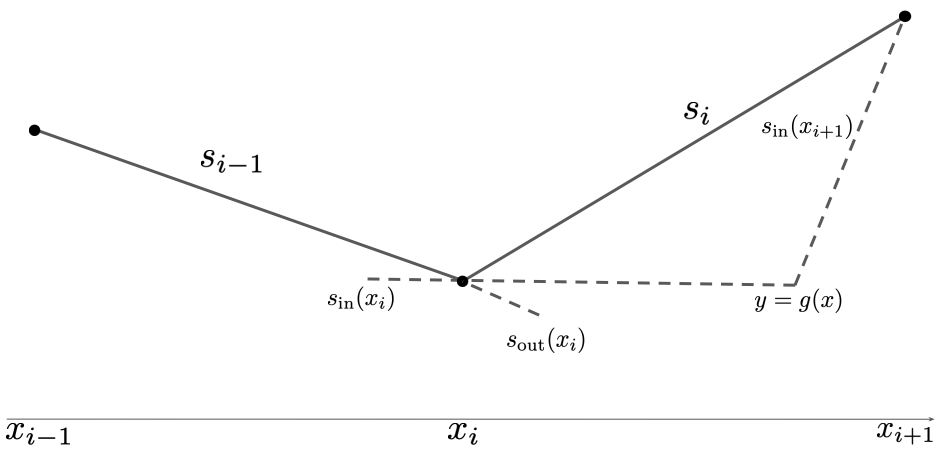

Figure 6: The function $g(x)$ used to derive a contradiction with the assumption that $s_{\text{out}}(x_i) < s_{\text{in}}(x_i)$ in Proposition A.7.

Claim A.8. *Suppose $a, b, c, d \in \mathbb{R}$ with $\max\{a, b\} \leq c < d$. Then for all $x \in \mathbb{R}$ we have*

$$|x - b| + 2d - a - b > |x - c| + c - a$$

Proof. Since $|x - c| \leq |x - d| + d - c$, we have

$$|x - b| + 2d - a - b - (|x - c| + c - a) \geq d - b > 0.$$

□

Next, again for the sake of contradiction, suppose that we still have $\epsilon_i = 1$ and $s_{i-1} \leq s_{\text{in}}(x_i) \leq s_{i+1}$ but also that $s_{\text{out}}(x_i) < s_{\text{in}}(x_i)$. Then, by Corollary A.3 we have $s_{\text{in}}(x_{i+1}) > s_i$. Moreover, by Proposition A.1 the total variation of Df on $(x_i - \delta, x_{i+1} + \delta)$ for all δ small enough is

$$|s_{\text{out}}(x_{i+1}) - s_{\text{in}}(x_{i+1})| + s_{\text{in}}(x_{i+1}) + s_{\text{in}}(x_i) - 2s_{\text{out}}(x_i).$$

Consider $g \in \text{PL(D)}$ defined to be equal to f for $x \in (x_i, x_{i+1})^c$ but for $x \in (x_i, x_{i+1})$ given by

$$g(x) = \max\left\{(x - x_i)s_{\text{in}}(x_i) + y_i,\ (x - x_{i-1})s_{\text{in}}(x_{i+1}) + y_{i+1}\right\}.$$

See Figure 6. The total variation of Dg on $(x_i - \epsilon, x_{i+1} + \epsilon)$ is

$$|s_{\text{out}}(x_{i+1}) - s_{\text{in}}(x_{i+1})| + s_{\text{in}}(x_{i+1}) - s_{\text{in}}(x_i).$$

Therefore the difference between the total variation of Df and Dg on $(x_i - \delta, x_{i+1} + \delta)$ is

$$2(s_{\text{in}}(x_i) - s_{\text{out}}(x_i)) > 0.$$

Since f and g agree on $(x_i, x_{i+1})^c$ this contradicts the minimality of $||Df||_{TV}$ in PL(D) and completes the proof of (15). □

Proposition A.7 allows us to translate information about the incoming slope $s_{\text{in}}(x_i)$ to outgoing information about $s_{\text{out}}(x_i)$. To make use of this, we also need a way to translate between outgoing information $s_{\text{out}}(x_i)$ and incoming information $s_{\text{in}}(x_{i+1})$. This is done in the following Proposition, whose conclusion is illustrated in Figure 7.

Proposition A.9 (How slope of f changes between x_i and x_{i+1} when ϵ_i, ϵ_{i+1} agree)**.** *If $\epsilon_i = 1$ and $s_{i-1} \leq s_{\text{in}}(x_i) \leq s_{\text{out}}(x_i) \leq s_i$, then*

$$\epsilon_{i+1} = 1 \quad \Longrightarrow \quad s_i \leq s_{\text{in}}(x_{i+1}) \leq s_{i+1} \tag{17}$$

Similarly, if $\epsilon_i = -1$ and $s_{i-1} \geq s_{\text{in}}(x_i) \geq s_{\text{out}}(x_i) \geq s_i$, then

$$\epsilon_{i+1} = -1 \quad \Longrightarrow \quad s_i \geq s_{\text{in}}(x_{i+1}) \geq s_{i+1} \tag{18}$$

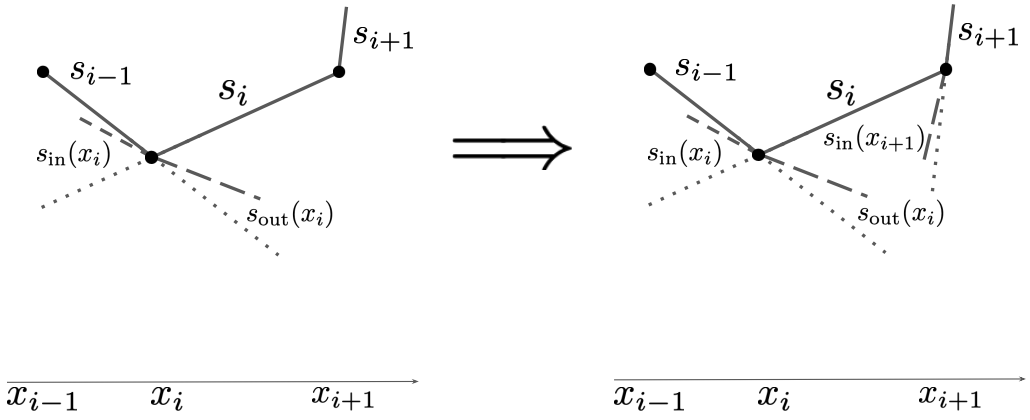

Figure 7: Illustration of the conclusion in Proposition A.9 when $\epsilon_i = \epsilon_{i+1} = 1$.

Proof. The relation (18) follows in the same way as (17), and so we focus on showing the latter. That is, we suppose $\epsilon_i = \epsilon_{i+1} = 1$ and that $s_{i-1} \leq s_{\text{in}}(x_i) \leq s_{\text{out}}(x_i) \leq s_i$. Corollary A.3 immediately gives $s_{\text{in}}(x_{i+1}) \geq s_i$. To complete the proof of (17) let us suppose for the sake of contradiction that in fact $s_{\text{in}}(x_{i+1}) > s_{i+1}$. To derive a contradiction, we need the following observation.

Lemma A.10. *Suppose that we have $\epsilon_{i+1} = 1$ and $s_{\text{in}}(x_{i+1}) > s_{i+1}$. Then we must have $s_{\text{out}}(x_{i+1}) < s_{\text{in}}(x_{i+1})$.*

Proof. If $i = m - 2$, then the conclusion follows immediately from the fact that by Proposition A.5 we have $s_{\text{out}}(x_{i+1}) = s_m$. If $i < m - 2$, let us suppose for the sake of contradiction that $s_{\text{out}}(x_{i+1}) \geq s_{\text{in}}(x_{i+1})$. In particular, we have $s_{\text{out}}(x_{i+1}) > s_{i+1}$. Hence, by Corollary A.3 we have

$$s_{\text{in}}(x_{i+2}) < s_{i+1}.$$

Also by Corollary A.3 since $s_{\text{in}}(x_{i+1}) > s_{i+1} > s_i$ we have

$$s_{\text{out}}(x_i) < s_i.$$

See Figure 8. The total variation of Df on $(x_i - \delta, x_{i+2} + \delta)$ for δ sufficiently small is therefore

$$|s_{\text{out}}(x_{i+2}) - s_{\text{in}}(x_{i+2})| + 2s_{\text{out}}(x_{i+1}) - s_{\text{in}}(x_{i+2}) - s_{\text{in}}(x_i)$$

Consider $g \in \text{PL(D)}$ that coincides with f on $(x_i, x_{i+2})^c$ and with $f_{\boldsymbol{D}}$ on (x_i, x_{i+2}). The total variation of Dg on $(x_i - \delta, x_{i+2} + \delta)$ for δ sufficiently small is

$$|s_{\text{out}}(x_{i+2}) - s_{i+1}| + s_{i+1} - s_{\text{in}}(x_i)$$

Using that $|s_{\text{out}}(x_{i+2}) - s_{i+1}| \leq |s_{\text{out}}(x_{i+2}) - s_{\text{in}}(x_{i+2})| + s_{i+1} - s_{\text{in}}(x_{i+2})$, we conclude that the difference between the total variation of Df and Dg is bounded below by

$$2\left(s_{\text{out}}(x_{i+1}) - s_{i+1}\right) > 0,$$

contradicting the minimality of $||Df||_{TV}$. □

Returning now to the proof of (17), we continue to assume that $s_{i-1} \leq s_{\text{in}}(x_i) \leq s_{\text{out}}(x_i) \leq s_i$ and $s_{\text{in}}(x_{i+1}) > s_{i+1}$. The previous Lemma ensures that therefore

$$s_{\text{in}}(x_{i+1}) > s_* := \max\left\{s_{i+1},\, s_{\text{out}}(x_{i+1})\right\}.$$

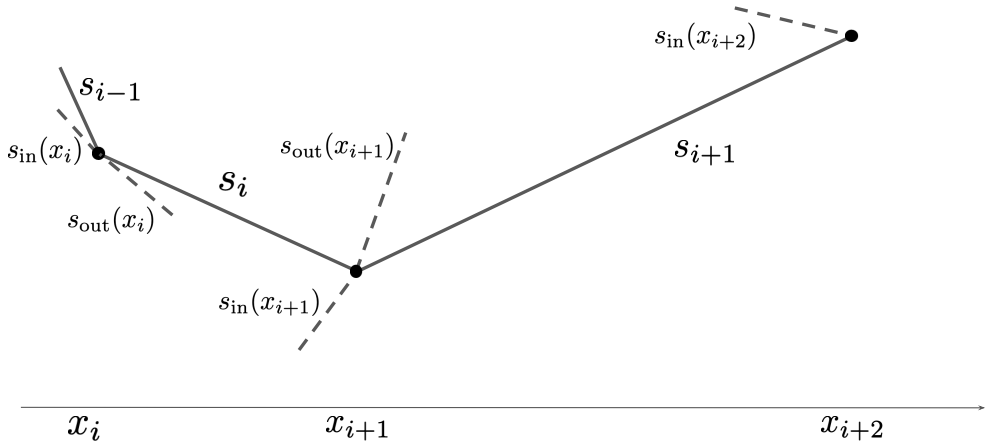

Figure 8: Illustration of hypotheses for contradiction in Lemma A.10.

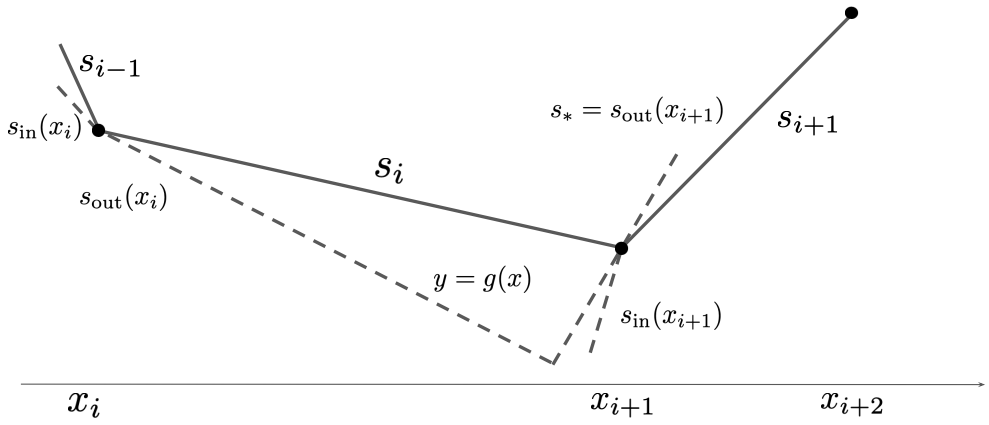

Figure 9: Illustration of the function g used for contradiction at the end of the proof of Proposition A.9.

From this last condition we conclude that the total variation of Df on $(x_i - \delta, x_{i+1} + \delta)$ for all δ sufficiently small is

$$2s_{\text{in}}(x_{i+1}) - s_{\text{in}}(x_i) - s_{\text{out}}(x_{i+1}).$$

Consider $g \in \text{PL(D)}$ defined to be equal to f on $(x_i, x_{i+1})^c$ but on (x_i, x_{i+1}) given by

$$g(x) = \max\left\{(x - x_{i+1})s_* + y_{i+1},\ (x - x_i)s_{\text{out}}(x_i) + y_i\right\}, \qquad x \in (x_i, x_{i+1}).$$

See Figure 9. Since $s_* \geq s_{i+1} > s_i$, we find that the total variation of Dg on $(x_i - \delta, x_{i+1} + \delta)$ for all δ small enough equals

$$s_* - s_{\text{in}}(x_i).$$

The difference of the total variation of Df and Dg on $(x_i - \delta, x_{i+1} + \delta)$ is therefore given by

$$s_{\text{in}}(x_{i+1}) - s_{i+1} + s_{\text{in}}(x_{i+1}) - s_* > 0.$$

This contradicts the minimality of $||Df||_{TV}$ among PL(D) and completes the proof of (17).

□

Proposition A.9 showed how to use information about the incoming and outgoing slopes of f at x_i to obtain information on the incoming slop at x_{i+1} if $\epsilon_i = \epsilon_{i+1}$. The following Proposition explains how to do this if instead $\epsilon_i \neq \epsilon_{i+1}$.

Proposition A.11 (How slope of f changes between x_i and x_{i+1} when ϵ_i, ϵ_{i+1} disagree)**.** *If $\epsilon_i = 1$ and $s_{i-1} \leq s_{\text{in}}(x_i) \leq s_{\text{out}}(x_i) \leq s_i$, then*

$$\epsilon_{i+1} = -1 \qquad \Longrightarrow \qquad s_{\text{out}}(x_i) = s_{\text{in}}(x_{i+1}) = s_i. \tag{19}$$

Similarly, if $\epsilon_i = -1$ and $s_{i-1} \geq s_{\text{in}}(x_i) \geq s_{\text{out}}(x_i) \geq s_i$, then

$$\epsilon_{i+1} = 1 \qquad \Longrightarrow \qquad s_{\text{out}}(x_i) = s_{\text{in}}(x_{i+1}) = s_i. \tag{20}$$

Proof. Relations (19) and (20) are proved in the same way, and so we focus on the former. To show (19), we suppose $\epsilon_i = 1$, $\epsilon_{i+1} = -1$ and that $s_{i-1} \leq s_{\text{in}}(x_i) \leq s_{\text{out}}(x_i) \leq s_i$. Suppose for the sake of contradiction that $s_{\text{out}}(x_i) < s_i$. Then, by Corollary A.3 we have $s_{\text{in}}(x_{i+1}) > s_i$. To see why this cannot occur, we give somewhat different arguments depending on whether $s_{\text{out}}(x_{i+1}) > s_i$ or $s_{\text{out}}(x_{i+1}) \leq s_i$.

Let us first suppose $s_{\text{out}}(x_{i+1}) > s_i$. By Corollary A.3 we have $s_{\text{in}}(x_{i+2}) < s_{i+1}$. Thus, the total variation of Df on $(x_i - \delta, x_{i+2} + \delta)$ equals

$$|s_{\text{out}}(x_{i+2}) - s_{\text{in}}(x_{i+2})| + s_{\text{out}}(x_{i+2}) - s_{\text{in}}(x_{i+2}) + |s_{\text{out}}(x_{i+1}) - s_{\text{in}}(x_{i+1})| + s_{\text{in}}(x_{i+1}) - s_{\text{in}}(x_i),$$

which is bounded below by

$$|s_{\text{out}}(x_{i+2}) - s_{\text{in}}(x_{i+2})| + 2s_{\text{out}}(x_{i+1}) - s_{\text{in}}(x_{i+2}) - s_{\text{in}}(x_i).$$

Define $g \in \text{PL(D)}$ to coincide with f on $(x_i, x_{i+2})^c$ and with $f_{\boldsymbol{D}}$ on (x_i, x_{i+2}). The total variation of Dg on $(x_i - \delta, x_{i+2} + \delta)$ is

$$|s_{\text{out}}(x_{i+2}) - s_{\text{in}}(x_{i+2})| + 2s_i - s_{\text{in}}(x_{i+2}) - s_{\text{in}}(x_i).$$

Hence, the difference between the total variation of Df and Dg is bounded below by

$$2(s_{\text{out}}(x_{i+1}) - s_i) > 0.$$

This contradicts the minimality of $||Df||_{TV}$. Let us now consider the other case: $s_{\text{out}}(x_{i+1}) \leq s_i$. In this case, we have that $s_{\text{in}}(x_{i+1}) > s_{\text{out}}(x_{i+1})$. Thus, the total variation of Df on $(x_i - \delta, x_{i+1} + \delta)$ is

$$2s_{\text{in}}(x_{i+1}) - s_{\text{in}}(x_i) - s_{\text{out}}(x_{i+1}).$$

Define $g \in \text{PL(D)}$ to coincide with f on $(x_i, x_{i+1})^c$ and with $f_{\boldsymbol{D}}$ on (x_i, x_{i+1}). The total variation of Dg on $(x_i - \delta, x_{i+1} + \delta)$ is

$$2s_i - s_{\text{out}}(x_{i+1}) - s_{\text{in}}(x_i).$$

Hence, the difference between the total variation of Df and Dg is bounded below by

$$2(s_{\text{in}}(x_{i+1}) - s_i) > 0.$$

This contradicts the minimality of $||Df||_{TV}$, completing the proof of Proposition A.11. □

We are now ready to show that any $f \in \text{RidgelessReLU}(\boldsymbol{D})$ satisfies (1) and (2). We already know from Propositions A.4 and A.5 that f satisfies properties (1a) and (1b). In order to check that f satisfies (1c) and (2), we will use the following result.

Lemma A.12. *Suppose $f \in \text{RidgelessReLU}(\boldsymbol{D})$. For $i = 2, \ldots, m-1$ we have*

$$\begin{aligned} \epsilon_i = 1 \qquad &\Longrightarrow \qquad s_{i-1} \leq s_{\text{in}}(x_i) \leq s_{\text{out}}(x_i) \leq s_i \\ \epsilon_i = -1 \qquad &\Longrightarrow \qquad s_{i-1} \geq s_{\text{in}}(x_i) \geq s_{\text{out}}(x_i) \geq s_i \end{aligned}$$

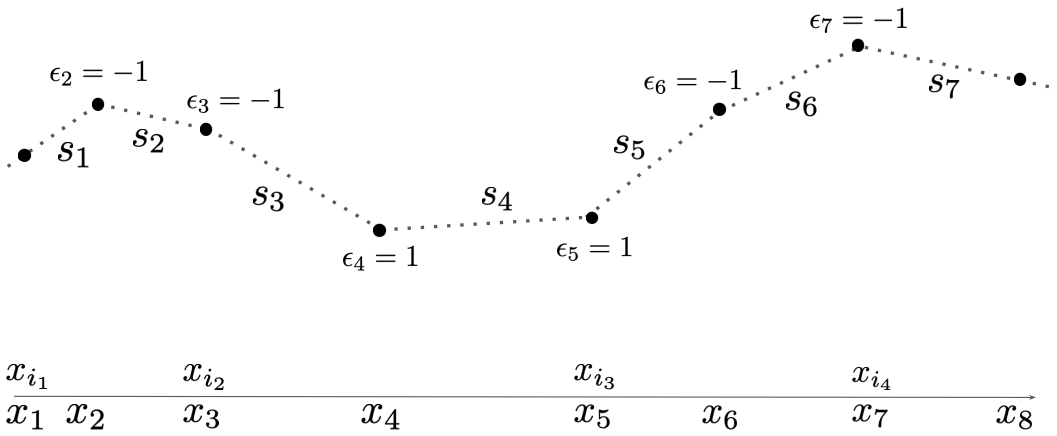

Figure 10: Illustration of the set of discrete inflections points $\boldsymbol{I}$ used in Proposition A.13.

Proof. We induct on i. When $i = 2$, we have from Proposition A.5 that

$$s_1 = s_{\text{in}}(x_2).$$

If $\epsilon_2 = 1$, we may therefore apply Proposition A.7 to conclude that $s_1 \leq s_{\text{in}}(x_1) \leq s_{\text{out}}(x_2) \leq s_2$, as desired. The case $\epsilon_2 = -1$ is similar, completing the base case. Let us now suppose we have the claim for $2, \ldots, i$. Suppose that $\epsilon_{i+1} = 1$ (the case $\epsilon_{i+1} = -1$ is similar). If $\epsilon_i \neq 1$, then we conclude from the definition of $\epsilon_{i+1} = 1$, the inductive hypothesis, and Propositions A.4 and A.11 that

$$s_i = s_{\text{in}}(x_{i+1}) \leq s_{i+1}$$

Hence, we may apply Proposition A.7 to conclude that $s_i = s_{\text{in}}(x_{i+1}) \leq s_{\text{out}}(x_{i+1}) \leq s_{i+1}$, as desired. This completes the inductive step and hence the proof of this Lemma. □

Lemma A.12 in combination with Corollary A.3 immediately implies that f satisfies property (2). Finally, in combination with Proposition A.11, Lemma A.12 also shows that f satisfies property (1c). This completes the proof that $f \in \text{RidgelessReLU}(\boldsymbol{D})$ satisfies properties (1) and (2). It remains to show that every f which satisfies Properties (1) and (2) belongs to $\text{RidgelessReLU}(\boldsymbol{D})$, which we now establish.

Proposition A.13. *Suppose $f \in \text{PL(D)}$ satisfies conditions (1) and (2) of Theorem 3.1. Then, f belongs to* $\text{RidgelessReLU}(\boldsymbol{D})$.

Proof. Define the set $\boldsymbol{I} \subseteq \{1, \ldots, m\}$ of discrete inflection points for the connect-the-dots interpolant $f_{\boldsymbol{D}}$ (see Figure 10):

$$\boldsymbol{I} := \{i \in \{2, \ldots, m-2\} \mid \epsilon_i \neq \epsilon_{i+1}\} \cup \{1, m-1\} = \left\{i_1 = 1 < i_2 < \cdots < i_{|\boldsymbol{I}|-1} < i_{|\boldsymbol{I}|} = m-1\right\}.$$

By construction, for each $q = 1, \ldots, |\boldsymbol{I}| - 1$ on the intervals $(x_{i_q}, x_{i_q+1}), \ldots, (x_{i_{q+1}}, x_{i_{q+1}+1})$ the sequence of slopes $s_{i_q}, \ldots, s_{i_{q+1}}$ of $f_{\boldsymbol{D}}$ is either non-increasing or non-decreasing. Hence,

$$\sum_{j=i_q}^{i_{q+1}-1} |s_j - s_{j+1}| = \left|s_{i_q} - s_{i_{q+1}}\right|$$

and we find

$$||Df_{\boldsymbol{D}}||_{TV} = \sum_{i=1}^{m-1} |s_i - s_{i+1}| = \sum_{q=2}^{|\boldsymbol{I}|} \left|s_{i_q} - s_{i_{q-1}}\right|. \tag{21}$$

The key observation is

$$f \in \mathrm{PL(D)} \text{ satisfies (1) and (2)} \qquad \Longrightarrow \qquad ||Df||_{TV} = ||Df_{\boldsymbol{D}}||_{TV} = \sum_{q=2}^{|\boldsymbol{I}|} \left| s_{i_q} - s_{i_{q-1}} \right|. \quad (22)$$

Indeed, by property (2), the function f is either convex or concave on any interval of the form $(x_{i_q}, x_{i_{q+1}+1})$. Therefore, Df is monotone on any such interval. Thus, we find that

$$||Df||_{TV} = \sum_{q=2}^{|\boldsymbol{I}|} \left| s_{\mathrm{out}}(f, x_{i_q}) - s_{\mathrm{out}}(f, x_{i_{q-1}}) \right|.$$

But property (1) guarantees that

$$s_{\mathrm{out}}(f, x_{i_q}) = s_{i_q}$$

and for all $q = 1, \ldots, |\boldsymbol{I}|$, proving (22). The proof of Proposition A.13 therefore follows from the following result, which was already observed in Theorem 3.3 of Savarese et al. (2019).

Lemma A.14. *We have*

$$\mathrm{RidgelessReLU}(\boldsymbol{D}) = \left\{ f \in \mathrm{PL}(\boldsymbol{D}) \;\middle|\; ||Df||_{TV} = ||Df_{\boldsymbol{D}}||_{TV} \right\} \quad (23)$$

Proof. Consider any $f \in \mathrm{RidgelessReLU}(\boldsymbol{D})$. We seek to show that $||Df||_{TV} \geq ||Df_{\boldsymbol{D}}||_{TV}$. Note that for any sequence of points $\xi_1 < \cdots < \xi_k$ at which $Df(\xi_j)$ exists, we have

$$||Df||_{TV} \geq \sum_{j=1}^{k-1} |Df(\xi_{j+1}) - Df(\xi_j)|.$$

We will now exhibit a set of points where the right hand side equals $||Df_{\boldsymbol{D}}||_{TV}$. To begin, note that by Proposition A.5 we have $f(x) = f_{\boldsymbol{D}}(x)$ for $x < x_2$ and $x > x_{m-1}$. For all $\xi_{i_1} \in (x_1, x_2) = (x_{i_1}, x_{i_1+1})$ and $\xi_{i_{|\boldsymbol{I}|}} \in (x_{m-1}, x_m) = (x_{i_{|\boldsymbol{I}|-1}}, x_{i_{|\boldsymbol{I}|}})$ we thus have

$$Df(\xi_{i_1}) = s_1, \qquad Df(\xi_{i_{|\boldsymbol{I}|}}) = s_m.$$

Further, for any $i = 2, \ldots, m-1$ on any interval (x_i, x_{i+1}), there exist $\xi_{i,\pm}$ such that $Df(\xi_{i,\pm})$ exist and

$$Df(\xi_{i,+}) \geq s_i, \qquad Df(\xi_{i,-}) \leq s_i.$$

In particular, for $q = 2, \ldots, |\boldsymbol{I}| - 1$ we may find ξ_{i_q} satisfying

$$\xi_{i_q} \in (x_{i_q}, x_{i_q+1}), \quad \mathrm{sgn}(s_{i_q} - Df(\xi_{i_q})) = \epsilon_{i_{q+1}}.$$

As we saw just before this Lemma, for each $i = 1, \ldots, |\boldsymbol{I}| - 1$ we have

$$\mathrm{sgn}\left(s_{i_{q+1}} - s_{i_q}\right) = \epsilon_{i_{q+1}}.$$

Hence, for each $q = 1, \ldots, |\boldsymbol{I}| - 1$ we conclude

$$\left| Df(\xi_{i_q}) - Df(\xi_{i_{q+1}}) \right| \geq \left| s_{i_q} - s_{i_{q+1}} \right|.$$

Thus,

$$||Df||_{TV} \geq \sum_{q=1}^{|\boldsymbol{I}|-1} \left| s_{i_q} - s_{i_{q+1}} \right| = ||Df_{\boldsymbol{D}}||_{TV},$$

as desired. □

□