Supplementary Material

In this document we give proofs for propositions (1) and (2) in the main paper. We use a slightly different notation for simplicity. We give a constructive proof for Proposition (2) that inherently implies Proposition (1). In the following section we give the necessary definitions and define the proximal operator for $\ell_\infty^T$-norm followed by proof in the next section.

1 Definitions

Let us consider a tree-structured set of groups of variables $G$, which are subsets of $\{1, \ldots, p\}$. The tree-structure definition follows [1], where two groups $g$ and $g'$ are either disjoint or one is included in the other.

Definition 1 (Tree-structured set of groups).
A set of groups $G \triangleq \{g\}_{g \in G}$ is said to be tree-structured in $\{1, \ldots, p\}$, if $\bigcup_{g \in G} g = \{1, \ldots, p\}$ and if for all $g, h \in G$, $g \cap h = \emptyset$, or $g \subseteq h$, or $h \subseteq g$. We also define for each group $g$,

- the set of variables $\text{root}(g) \subseteq g$ is such that $i \in \text{root}(g)$ is not in $g'$ for all group $g' \subseteq g$;
- the set of groups $\text{children}(g)$ is the set of groups $g'$ such that $g' \subseteq g$.

We are now interested in the following optimization problem

$$
\min_{w \in \mathbb{R}^p} \frac{1}{2} \|u - w\|_2^2 + \lambda \sum_{g \in G} \|w_g\|_\infty.
$$

(1)

Following [1], it can be solved by Algorithm [1] where $\Pi_\lambda$ is the Euclidean projection on the $\ell_1$-ball of radius $\lambda$.

Lemma 1 (Equivalent Views of the $\ell_\infty$-proximal Operator).
Let us consider the proximal operator $\text{Prox}_\lambda^T$:

$$
\text{Prox}_\lambda^T: u \mapsto \arg \min_{w \in \mathbb{R}^p} \frac{1}{2} \|u - w\|_2^2 + \lambda \|w\|_\infty.
$$

Then,

$$
[\text{Prox}_\lambda^T(u)]_g = u_g - \Pi_\lambda(u_g),
$$

(2)
Algorithm 1 Computation of the Proximal Operator.

Inputs: $u \in \mathbb{R}^p$ and an ordered tree-structured set of groups $G$ with root $g_0$.
Initialization: $w \leftarrow u$;
Call recursiveProx$(g_0)$;
Return $w$.

Procedure recursiveProx$(g)$

1: for $h \in \text{child}(g)$ do
2: Call recursiveProx$(h)$;
3: end for
4: $w_g \leftarrow w_g - \Pi_{\lambda}(w_g)$.

and there exists $\tau \geq 0$ such that for all $j \in g$,

$$|\text{Prox}^g_\lambda(u)|_j = \text{sign}(u_j) \min(|u_j|, \tau) \quad \text{and} \quad (3)$$

$$\left\{ \begin{array}{l}
\|\Pi_{\lambda}(u_g)\|_1 = \sum_{j \in g} \max(|u_j| - \tau, 0) = \lambda \quad \text{or} \quad \tau = 0
\end{array} \right\}. \quad (4)$$

Proof. The proof of Eq. (2) can be found in [1]. The proof of Eq. (4) consists of noticing that the projection on the $\ell_1$-ball is obtained by a soft-thresholding operator [1]. In other words, there exists $\tau \geq 0$ such that $[\Pi_{\lambda}(u)]_j = \text{sign}(u_j) \max(|u_j| - \tau, 0)$ for all $j$ in $g$. We notice that by definition of the Euclidean projection, either $\|\Pi_{\lambda}(u_g)\|_1 < \lambda$ and $\Pi_{\lambda}(u_g) = u_g$ (meaning $\tau = 0$), or $\|\Pi_{\lambda}(u_g)\|_1 = \lambda$. This yields (4).

By using the definition of $\text{Prox}^g_\lambda$, we see that Algorithm 1 in fact performs a composition of proximal operators. Suppose that the groups in $G = \{g_1, \ldots, g_k\}$ are ordered according to depth-first search order, we have

$$\text{Prox}_{\lambda \Omega} = \text{Prox}^{g_k} \circ \ldots \circ \text{Prox}^{g_1},$$

where $\Omega$ is the tree-structured penalty $\Omega(w) = \sum_{g \in G} \|w_g\|_{\infty}$, and $\circ$ is a composition operator.

We now have the following (Proposition 2 of main paper) to compose proximal step over constant value non-branching paths or nested groups. We prove this by showing that in consecutive projections the $\tau$ in (3) can only be smaller than the previous one forcing the values along a non-branching path to be equal.

Lemma 2 (Composition Lemma Along Nested Groups). Assume that for all groups $g$ in $G$, root$(g)$ is a singleton $\{r(g)\}$. Consider a particular group $g$ with a single child $g'$, such that $u_{r(g)} = u_{r(g')}$. Then,

$$\left(\text{Prox}^{g}_\lambda \circ \text{Prox}^{g'}_\lambda\right)(u) = \text{Prox}^{g}_2(\lambda)(u).$$

2
Since \( \tau \) define \( g \) since all the proximal operators only affect the variables in \( g \). We also have \( \text{assign the stored signs to the result} \ [1] \). We also have that
\[
\left[ (\text{Prox}_\lambda \circ \text{Prox}_{\lambda'}^g) (u) \right]_j = [\text{Prox}_{2\lambda}^g (u)]_j = u_j \quad \text{for all } j \not\in g,
\]
since all the proximal operators only affect the variables in \( g \) and \( g' \). Let us now define \( v \equiv \text{Prox}_{\lambda}^g (u), \ w^* \equiv \text{Prox}_{\lambda}^g (v) \).

Consider \( \tau' \) defined in Lemma 1 such that \( v_{g'} = \min(u_{g'}, \tau') \), and \( \tau \) such that \( w^*_g = \min(v_g, \tau) \).

**First step:** \( \tau \leq \tau' \):
Let us proceed by contradiction and assume that \( \tau < \tau' \). Then, we have \( v_{g'} \leq \tau \) and thus, Eq. (4) applied to the group \( g \) gives us that \( u_{r(g)} - \tau = v_{r(g)} - \tau = \lambda \)
since \( \tau \neq 0 \) and \( g = g' \cup \{r(g)\} \). Note also that \( u_{r(g')} - \tau' \leq ||\Pi_{\lambda}(u_{g'})||_1 \leq \lambda \)
according to Eq. (4) applied to the group \( g' \). Since \( u_{r(g')} = u_{r(g)} \), we have \( u_{r(g')} - \tau' \leq u_{r(g)} - \tau, \) and \( \tau \leq \tau' \), which is a contradiction.

**End of the proof:**
By using Eq. (4), and using the fact that \( \tau \leq \tau' \), we now have a closed form solution for \( w^*_g \):
\[
w^*_g = \min(u_g, \tau).
\]

We now consider two cases

- if \( \tau = 0 \), we have \( w^*_g = 0 \), and thus \( v_g = \Pi_{\lambda}(v_g) \), meaning that \( \|v_g\|_1 \leq \lambda \).
  Thus, \( \|u_g\|_1 = \|v_g\|_1 + \|u_{g'} - v_{g'}\|_1 \leq \lambda + \|\Pi_{\lambda}(u_{g'})\|_1 \leq 2\lambda \).
  Thus, \( \text{Prox}_{2\lambda}^g (u) \). Let us now apply again Eq. (4).
- if \( \tau > 0 \), we define the quantity \( z_g = u_g - w^*_g = \max(u_g - \tau, 0) \), which has the form of an orthogonal projection of \( u_g \) onto the \( \ell_1 \)-ball of some radius \( \lambda' \) (see [1]). It remains to compute \( \|z_g\|_1 \) to know the radius of \( \lambda' \). We have
\[
\|z_g\|_1 = \|u_g - w^*_g\|_1 = \|u_g - v_g + v_g - w^*_g\|_1 = \|u_{g'} - v_{g'}\|_1 + \|v_g - w^*_g\|_1 = 2\lambda,
\]
where we apply again Eq. (4). Thus, \( z_g = \Pi_{2\lambda}(u_g) \) and \( w^*_g = \text{Prox}_{2\lambda}^g (u) \).

This proof can be put together for paths with more than two nested groups to inductively construct single-step proximal projections for longer paths.

It is easy to see from this definition [4] that all entries with the same value \( u_j = \delta \forall j \) will continue to share a value after applying the proximal operator \( \min(\delta, \tau) \). We see from [2] that all entries at nested groups will be projected to the same value. This in fact turns out to be a single projection with the \( \lambda \) scaled appropriately.

These two put together we have the property that constant value non-branching paths are preserved.
References