Algorithm 1 Hierarchical Proximal Operator

procedure HIER-PROX($v, u; \lambda, \bar{\lambda}, M$)
  for $s \in \{0, \ldots, K\}$ **do**
    Sort the entries of u_s into $|u_{(s,1)}| \geq \ldots \geq |u_{(s,K)}|$
    $a_s := \lambda - M \sum_{i=1}^{s} (|u_{(i)}| - \bar{\lambda})$
    $x := \frac{1}{1 + sM^2} \left(1 - \frac{a_s}{\|v\|_2}\right)$
    $w := M \cdot x \cdot \|v\|_2$
    Find $\tilde{i}$, the first $i \in \{0, \ldots, K\}$ such that $|w_{(j,i+1)}| \leq w_i \leq |w_{(j,i)}|$
    $\beta_s^* \leftarrow x_{\tilde{i}} \cdot v$
    $\theta_s^* \leftarrow \text{sign}(u) \cdot \min(\mathcal{S}_{\bar{\lambda}}(u), w_{\tilde{i}})$
  end for
  return (β^*, θ^*)
end procedure
Notation: d denotes the number of features; K denotes the size of the first hidden layer.