A Appendix

A.1 Straight-through Estimator

The Straight-Through Estimator (STE) [2] is a commonly used technique in deep learning paradigms that incorporate discrete or non-differentiable functions, for example, binarization. This method facilitates the approximation of gradients through these operations, thereby enabling end-to-end model training involving non-differentiable components. Consequently, we utilize STE for backpropagation and approximate its gradient $G'(x)$ correspondingly.

$$G'(x) \approx 1. \tag{5}$$

A.2 Proofs for Lemmas

Lemma 1.

$$\mathbb{E}_x\left[\mathbf{1}_{\{\theta_1\cdot x>\theta_2\cdot x\}}\mathbf{1}_{\{\hat{\theta}_1\cdot x>\hat{\theta}_2\cdot x\}}\right] = \frac{1}{2}\left(1-\frac{\arccos\left(\frac{(\theta_1-\theta_2)\cdot(\hat{\theta}_1-\hat{\theta}_2)}{\|\theta_1-\theta_2\|_2\|\hat{\theta}_1-\hat{\theta}_2\|_2}\right)}{\pi}\right).$$

Proof. Consider the plane spanned by vector $\theta_1-\theta_2$ and $\hat{\theta}_1-\hat{\theta}_2$ and the projection of x to this plane, the two indicator function requires the angle $< Px, \theta_1-\theta_2 >$ and angle $< Px, \hat{\theta}_1-\hat{\theta}_2 >$ to be smaller than $\frac{\pi}{2}$. Evaluating the expectation over $\mathcal{X}$ is equivalent to evaluating the intersection region of two semicircles. Therefore the result is $\frac{\pi-\arccos\left(\frac{(\theta_1-\theta_2)\cdot(\hat{\theta}_1-\hat{\theta}_2)}{\|\theta_1-\theta_2\|_2\|\hat{\theta}_1-\hat{\theta}_2\|_2}\right)}{2\pi}$.

Lemma 2. *Let $\Delta\theta=\theta_1-\theta_2, \Delta\hat{\theta}=\hat{\theta}_1-\hat{\theta}_2$. When the coordinates of vector $\Delta\theta$ are ordered by absolute value: $1 \geq |\Delta\theta_1| \geq |\Delta\theta_2| \geq \cdots \geq |\Delta\theta_d|$. Then we have the following equality:*

$$\sup_{\Delta\hat{\theta}\in\{-1,0,1\}^d} \frac{\Delta\theta\cdot\Delta\hat{\theta}}{\|\Delta\theta\|_2\|\Delta\hat{\theta}\|_2} = \sup_{1\leq j\leq d}\left\{\frac{\sum_{i=1}^{j}|\Delta\theta_j|}{\sqrt{j}\|\Delta\theta\|_2}\right\}.$$

Proof. By the definition of the supremum, iterate over the list $\Delta\hat{\theta} \in [\mathbf{e}_1, \mathbf{e}_1+\mathbf{e}_2, \ldots, \mathbf{e}_1+\mathbf{e}_2+\cdots+\mathbf{e}_d]$, $\mathbf{e}_i$ is the unit vector with the same sign as $\Delta\theta_i$, we know

$$\sup_{\Delta\hat{\theta}\in\{-1,0,1\}^d} \frac{\Delta\theta\cdot\Delta\hat{\theta}}{\|\Delta\theta\|_2\|\Delta\hat{\theta}\|_2} \geq \sup_{1\leq j\leq d}\left\{\frac{\sum_{i=1}^{j}|\Delta\theta_j|}{\sqrt{j}\|\Delta\theta\|_2}\right\}.$$

Now we show the $\leq$ part. We show that when the $\Delta\theta$'s coordinates are ordered, the optimal $\Delta\hat{\theta}$ is of the form

$$(\mathrm{sign}(\Delta\theta_1), \ldots, \mathrm{sign}(\Delta\theta_j), 0, \ldots, 0).$$

For any $\Delta\hat{\theta}$ with norm $\sqrt{j}$,

$$\Delta\theta\cdot\Delta\hat{\theta} \leq \sum_{i=1}^{j}|\Delta\theta_j|.$$

Therefore,

$$\sup_{\Delta\hat{\theta}\in\{-1,0,1\}^d} \frac{\Delta\theta\cdot\Delta\hat{\theta}}{\|\Delta\theta\|_2\|\Delta\hat{\theta}\|_2} = \sup_j \sup_{|\Delta\hat{\theta}|=\sqrt{j},} \frac{\Delta\theta\cdot\Delta\hat{\theta}}{\|\Delta\theta\|_2\|\Delta\hat{\theta}\|_2} \le \sup_{1\le j\le d}\{\frac{\sum_{i=1}^{j}|\Delta\theta_j|}{\sqrt{j}\|\Delta\theta\|_2}\}.$$

Lemma 3.

$$\inf_{\theta_1,\theta_2}\sup_{\hat{\theta}_1,\hat{\theta}_2}\left[1-\frac{\arccos(\frac{(\theta_1-\theta_2)\cdot(\hat{\theta}_1-\hat{\theta}_2)}{\|\theta_1-\theta_2\|_2\|\hat{\theta}_1-\hat{\theta}_2\|_2})}{\pi}\right] \le 1-\frac{\arccos(\frac{1}{\sqrt{\sum_{j=1}^{d}(\sqrt{j}-\sqrt{j-1})^2}})}{\pi}.$$

Proof. We will show the $\le$ part by construction. Set $\theta_1 = (1, \sqrt{2}-\sqrt{1}, \ldots, \sqrt{d}-\sqrt{d-1})$, $\theta_2 = (0, 0, \ldots, 0)$. According to the Lemma 2 and the monotonicity of arccos function, we have

$$\begin{aligned}\inf_{\theta_1,\theta_2}\sup_{\hat{\theta}_1,\hat{\theta}_2}\left[1-\frac{\arccos(\frac{(\theta_1-\theta_2)\cdot(\hat{\theta}_1-\hat{\theta}_2)}{\|\theta_1-\theta_2\|_2\|\hat{\theta}_1-\hat{\theta}_2\|_2})}{\pi}\right] &\le \sup_{\hat{\theta}_1,\hat{\theta}_2}\left[1-\frac{\arccos(\frac{(\theta_1-\theta_2)\cdot(\hat{\theta}_1-\hat{\theta}_2)}{\|\theta_1-\theta_2\|_2\|\hat{\theta}_1-\hat{\theta}_2\|_2})}{\pi}\right]\\ &= 1-\frac{\arccos(\sup_{\hat{\theta}_1,\hat{\theta}_2}\frac{\theta_1-\theta_2)\cdot(\hat{\theta}_1-\hat{\theta}_2)}{|\theta_1-\theta_2||\hat{\theta}_1-\hat{\theta}_2)|})}{\pi}\\ &= 1-\frac{\arccos(\frac{1}{\sqrt{\sum_{j=1}^{d}(\sqrt{j}-\sqrt{j-1})^2}})}{\pi}.\end{aligned}$$

Lemma 4.

$$\inf_{\theta_1,\theta_2}\sup_{\hat{\theta}_1,\hat{\theta}_2}\left[1-\frac{\arccos(\frac{(\theta_1-\theta_2)\cdot(\hat{\theta}_1-\hat{\theta}_2)}{\|\theta_1-\theta_2\|_2\|\hat{\theta}_1-\hat{\theta}_2\|_2})}{\pi}\right] \ge 1-\frac{\arccos(\frac{1}{\sqrt{\sum_{j=1}^{d}(\sqrt{j}-\sqrt{j-1})^2}})}{\pi}.$$

Proof. Proof by contradiction. Assume there exists θ_1^*, θ_2^* such that the LHS is smaller than $1-\frac{\arccos(\frac{1}{\sqrt{\sum_{j=1}^{d}(\sqrt{j}-\sqrt{j-1})^2}})}{\pi}$, by monotonicity of cosine function we know

$$C_0 := \sup_{\hat{\theta}_1,\hat{\theta}_2}\frac{(\theta_1^*-\theta_2^*)\cdot(\hat{\theta}_1-\hat{\theta}_2)}{|\theta_1^*-\theta_2^*||\,\hat{\theta}_1-\hat{\theta}_2\,|} < \frac{1}{\sqrt{\sum_{j=1}^{d}(\sqrt{j}-\sqrt{j-1})^2}} =: C_d.$$

Denote $\Delta\theta^* = \theta_1^*-\theta_2^*$. Without loss of generality, we assume $\Delta\theta^* \in [0,1]^d$, $\|\Delta\theta^*\|_2 = 1$ and

$$\Delta\theta_1^* \ge \Delta\theta_2^* \ge \cdots \ge \Delta\theta_d^*.$$

Starting from $\Delta\theta_1^*, \ldots, \Delta\theta_d^*$, we construct another feasible solution $\Delta\theta_1, \ldots, \Delta\theta_d$ without increasing the corresponding supremum value beyond C_0. However, if we compare $\Delta\theta_1, \ldots, \Delta\theta_d$ element-wisely with $(\sqrt{k}-\sqrt{k-1})C_0, 1 \le k \le d$, the first $\Delta\theta_k$ that is not equal to $(\sqrt{k}-\sqrt{k-1})C_0$ is greater than $(\sqrt{k}-\sqrt{k-1})C_0$, this gives us the contradiction to the C_0's definition.

Also notice $\sum_{i=1}^{d}(\Delta\theta_i^*)^2 = 1, C_0 < C_d$, there always exists index k satisfying $\Delta\theta_k^* > (\sqrt{k} - \sqrt{k-1})C_0$.

Assume the first θ_i that is not equal to $(\sqrt{i} - \sqrt{i-1})C_0$ is still smaller than $(\sqrt{i} - \sqrt{i-1})C_0$. By the above paragraph, we can find the first $\theta_k, k > i$ with $\theta_k^* > (\sqrt{k} - \sqrt{k-1})C_0$.

Then we adjust (θ_i^*, θ_k^*) to $((\sqrt{i} - \sqrt{i-1})C_0, \sqrt{(\theta_i^*)^2 + (\theta_k^*)^2 - (\sqrt{i} - \sqrt{i-1})^2 C_0^2})$. We can verify that the assumed inequalities continue to hold. (There are cases for $(\theta_i^*)^2 + (\theta_k^*)^2 - (\sqrt{i} - \sqrt{i-1})^2 C_0^2 \leq (\sqrt{k} - \sqrt{k-1})^2 C_0^2$, then we just end this modification with $(\sqrt{(\theta_i^*)^2 + (\theta_k^*)^2 - (\sqrt{k} - \sqrt{k-1})^2 C_0^2}, (\sqrt{k} - \sqrt{k-1})C_0)$ and then repeat the procedure.)

As the above procedure repeats, number $\#\{k|\theta_k = (\sqrt{k} - \sqrt{k-1})C_0\}$ is strictly increased. When it stopped, the first non-$(\sqrt{k} - \sqrt{k-1})C_0$ term is larger than the $(\sqrt{k} - \sqrt{k-1})C_0$ and this gives us the contradiction.