Supplementary Material

Implicit NNs are Almost Equivalent to Not-so-deep Explicit NNs for High-dimensional Gaussian Mixtures

A Preliminaries

We consider n data vectors $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n \in \mathbb{R}^p$ independently drawn from one of the K-class Gaussian mixture $\mathcal{C}_1, \cdots, \mathcal{C}_K$ and denote $\boldsymbol{X} = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n] \in \mathbb{R}^{p \times n}$, with class $\mathcal{C}_a$ having cardinality n_a, *i.e.*, for $\boldsymbol{x}_i \in \mathcal{C}_a$, we have

$$\boldsymbol{x}_i \sim \mathcal{N}(\boldsymbol{\mu}_a/\sqrt{p}, \boldsymbol{C}_a/p)$$

Assumption 4. *We assume that, as $n \to \infty$, we have, for $a \in \{1, \cdots, K\}$ that,*

- $p/n \to c \in (0, \infty)$ *and* $n_a/n \to c_a \in (0,1)$*; and*
- $\|\boldsymbol{\mu}_a\| = \mathcal{O}(1)$*; and*
- *for* $\boldsymbol{C}^\circ \equiv \sum_{a=1}^K \frac{n_a}{n} \boldsymbol{C}_a$ *and* $\boldsymbol{C}_a^\circ \equiv \boldsymbol{C}_a - \boldsymbol{C}^\circ$*, we have* $\|\boldsymbol{C}_a\| = \mathcal{O}(1)$, $\operatorname{tr} \boldsymbol{C}_a^\circ = \mathcal{O}\left(p^{\frac{1}{2}}\right)$ *and* $\operatorname{tr}(\boldsymbol{C}_a \boldsymbol{C}_b) = \mathcal{O}(p)$ *for* $a, b \in \{1, \cdots, K\}$*; and*
- $\tau_0 = \sqrt{\operatorname{tr} \boldsymbol{C}^\circ / p}$ *converges in* $(0, \infty)$.

Some quantities. We first introduce the following notations. For $\boldsymbol{x}_i, \boldsymbol{x}_j \in \mathbb{R}^p$ with $i \neq j$, let

$$\boldsymbol{x}_i = \boldsymbol{\mu}_i/\sqrt{p} + \boldsymbol{z}_i/\sqrt{p}, \quad \boldsymbol{x}_j = \boldsymbol{\mu}_j/\sqrt{p} + \boldsymbol{z}_j/\sqrt{p},$$

so that $\boldsymbol{z}_i \sim \mathcal{N}(0, \boldsymbol{C}_i)$, $\boldsymbol{z}_j \sim \mathcal{N}(0, \boldsymbol{C}_j)$, and

$$\begin{aligned}
\boldsymbol{x}_i^\top \boldsymbol{x}_j &= \underbrace{\frac{1}{p} \boldsymbol{z}_i^\top \boldsymbol{z}_j}_{\mathcal{O}(p^{-1/2})} + \underbrace{\frac{1}{p} \boldsymbol{\mu}_i^\top \boldsymbol{\mu}_j + \frac{1}{p}(\boldsymbol{\mu}_i^\top \boldsymbol{z}_j + \boldsymbol{\mu}_j^\top \boldsymbol{z}_i)}_{\mathcal{O}(p^{-1})}, \\
\psi_i &= \frac{1}{p}\|\boldsymbol{z}_i\|^2 - \frac{1}{p} \operatorname{tr} \boldsymbol{C}_i = \mathcal{O}\left(p^{-1/2}\right), \quad s_i \equiv \|\mu_i\|^2/p + 2\boldsymbol{\mu}_i^\top \boldsymbol{z}_i/p = \mathcal{O}\left(p^{-1}\right) \\
t_i &= \frac{1}{p} \operatorname{tr} \boldsymbol{C}_i^\circ = \mathcal{O}\left(p^{-1/2}\right), \quad \tau_0 = \sqrt{\frac{1}{p} \boldsymbol{C}^\circ} = \mathcal{O}(1), \\
\chi_i &= \underbrace{t_i + \psi_i}_{\mathcal{O}(p^{-1/2})} + \underbrace{s_i}_{\mathcal{O}(p^{-1})} = \|\boldsymbol{x}_i\|^2 - \tau_0^2.
\end{aligned}$$

It can be checked that

$$\begin{aligned}
\|\boldsymbol{x}_i\|^2 &= \frac{1}{p}(\boldsymbol{\mu}_i + \boldsymbol{z}_i)^\top (\boldsymbol{\mu}_i + \boldsymbol{z}_i) = \frac{1}{p}\|\boldsymbol{\mu}_i\|^2 + \frac{2}{p} \boldsymbol{\mu}_i^\top \boldsymbol{z}_i + \frac{1}{p} \boldsymbol{z}_i^\top \boldsymbol{z}_i \\
&= \underbrace{\frac{1}{p}\|\boldsymbol{\mu}_i\|^2 + \frac{2}{p} \boldsymbol{\mu}_i^\top \boldsymbol{z}_i}_{\equiv s_i = \mathcal{O}(p^{-1})} + \underbrace{\frac{1}{p} \operatorname{tr} \boldsymbol{C}^\circ}_{\equiv \tau_0^2 = \mathcal{O}(1)} + \underbrace{\frac{1}{p} \operatorname{tr} \boldsymbol{C}_i^\circ}_{\equiv t_i = \mathcal{O}(p^{-1/2})} + \underbrace{\psi_i}_{\mathcal{O}(p^{-1/2})}
\end{aligned}$$

By Taylor-expanding $\sqrt{\|\boldsymbol{x}_i\|^2}$ around τ_0^2, we have

$$\|\boldsymbol{x}_i\| = \tau_0 + \frac{1}{2\tau_0}(\|\boldsymbol{\mu}_i\|^2/p + 2\boldsymbol{\mu}_i^\top \boldsymbol{z}_i/p + t_i + \psi_i) - \frac{1}{8\tau_0^3}(t_i + \psi_i)^2 + \mathcal{O}\left(p^{-3/2}\right). \tag{22}$$

Additionally, we denote S_{ij} terms of the form

$$S_{ij} \equiv S_{ij}(\gamma_1, \gamma_2) = \frac{1}{p} \boldsymbol{z}_i^\top \boldsymbol{z}_j (\gamma_1(t_i + \psi_i) + \gamma_2(t_j + \psi_j)),$$

for random or deterministic scalars $\gamma_1, \gamma_2 = \mathcal{O}(1)$ (with high probability when being random). Note that $S_{ij} = \mathcal{O}\big(p^{-1}\big)$ and more importantly, it leads to, in matrix form, a matrix of spectral norm order $\mathcal{O}\big(p^{-1}\big)$ Couillet and Benaych-Georges (2016).

Moreover, we introduce an important quantity τ_*, which plays a crucial role in our proof. We recursively define τ_l as

$$\tau_l = \sqrt{\sigma_a^2 \mathbb{E}[\phi^2(\tau_{l-1}\xi)] + (1-\sigma_a^2)\tau_0^2},$$

for $l = 1, 2, \cdots$.

Lemma A.1. *Let Condition 2 hold. As $l \to \infty$, τ_l converges to a fixed point τ^* such that*

$$\lim_{l\to\infty} \tau_l \equiv \tau^* = \sqrt{\sigma_a^2 \mathbb{E}[\phi^2(\tau^*\xi)] + (1-\sigma_a^2)\tau_0^2}.$$

Proof. Let $t = \tau_{l-1}^2$. By taking the derivative with respect to t on the RHS of Eq. (11), we have

$$\begin{aligned}
&\frac{\partial}{\partial t}\left(\sigma_a^2 \mathbb{E}\left[f(\tau_{l-1}\xi)\right] + (1-\sigma_a^2)\tau_0^2\right) \\
=&\sigma_a^2 \frac{\partial}{\partial t}\mathbb{E}\left[f(\sqrt{t}\cdot\xi)\right] \\
=&\sigma_a^2 \frac{\partial}{\partial t}\left(\int \frac{1}{\sqrt{2\pi}} f(\sqrt{t}\cdot x)e^{-\frac{x^2}{2}}dx\right) \\
=&\sigma_a^2 \frac{1}{\sqrt{2\pi}}\int f'(\sqrt{t}\cdot x)\frac{x}{2\sqrt{t}}e^{-\frac{x^2}{2}}dx \\
=&\frac{\sigma_a^2}{2}\cdot\mathbb{E}[f''(\tau_{l-1}\xi)], \quad \text{by the Gaussian integration by parts formula,}
\end{aligned}$$

which implies that the RHS of Eq. (11) is a *contractive mapping* if

$$\sigma_a^2 < \frac{2}{L_2}.$$

As a result, under Condition 2, the unique fixed point τ_* exists. $\square$

B Proof of Theorem 1

We prove Theorem 1 by performing induction on the hypothesis that $\|\boldsymbol{G}^{(l-1)} - \widetilde{\boldsymbol{G}}^{(l-1)}\| \to 0$ holds at layer $l-1$ with

$$\widetilde{\boldsymbol{G}}^{(l-1)} \equiv \alpha_{l-1,1}\boldsymbol{X}^\top\boldsymbol{X} + \boldsymbol{V}\boldsymbol{C}^{(l-1)}\boldsymbol{V}^\top + (\tau_{l-1}^2 - \tau_0^2\alpha_{l-1,1} - \tau_0^4\alpha_{l-1,3})\boldsymbol{I}_n,$$

for $\boldsymbol{C}^{(l-1)} = \begin{bmatrix} \alpha_{l-1,2}\boldsymbol{t}\boldsymbol{t}^\top + \alpha_{l-1,3}\boldsymbol{T} & \alpha_{l-1,2}\boldsymbol{t} \\ \alpha_{l-1,2}\boldsymbol{t}^\top & \alpha_{l-1,2}\end{bmatrix}$, and work on the CK matrix $\boldsymbol{G}^{(l)}$ at layer l.

The following lemma plays an important role in our proof.

Lemma B.1 (Gu et al. (2022))**.** *Assume that the activation function $\phi(\cdot)$ is "centred", such that $\mathbb{E}[\phi(\tau\xi)] = 0$, and*

$$\boldsymbol{G}_{ii}^{(l-1)} = \tau_{l-1}^2 + \alpha_{l-1,4}\chi_i + \alpha_{l-1,5}(t_i+\psi_i)^2 + \mathcal{O}\Big(p^{-3/2}\Big)$$

$$\boldsymbol{G}_{ij}^{(l-1)} = \alpha_{l-1,1}\boldsymbol{x}_i^\top\boldsymbol{x}_j + \alpha_{l-1,2}(t_i+\psi_i)(t_j+\psi_j) + \alpha_{l-1,3}\left(\frac{1}{p}\boldsymbol{z}_i^\top\boldsymbol{z}_j\right)^2 + S_{ij} + \mathcal{O}\Big(p^{-3/2}\Big).$$

It holds that

$$\begin{aligned}
&\mathbb{E}\left[\phi\left(\sqrt{\boldsymbol{G}_{ii}^{(l-1)}}\cdot\xi_i\right)\times\phi\left(\frac{\boldsymbol{G}_{ij}^{(l-1)}}{\sqrt{\boldsymbol{G}_{ii}^{(l-1)}}}\cdot\xi_i + \sqrt{\boldsymbol{G}_{jj}^{(l-1)} - \frac{\left(\boldsymbol{G}_{ij}^{(l-1)}\right)^2}{\boldsymbol{G}_{ii}^{(l-1)}}}\cdot\xi_j\right)\right] \\
=&\mathbb{E}[\phi'(\tau_{l-1}\xi)]^2\boldsymbol{G}_{ij}^{(l-1)} + \frac{\alpha_{l-1,1}^2}{2}\mathbb{E}[\phi''(\tau_{l-1}\xi)]^2\left(\frac{1}{p}\boldsymbol{z}_i^\top\boldsymbol{z}_j\right)^2 + \frac{\alpha_{l-1,4}^2}{4}\mathbb{E}[\phi''(\tau_{l-1}\xi)]^2(t_i+\psi_i)(t_j+\psi_j) \\
&+ S_{ij} + \mathcal{O}\Big(p^{-3/2}\Big).
\end{aligned}$$

On the diagonal. By induction hypothesis on the layer $l-1$, we have

$$\boldsymbol{G}_{ii}^{(l-1)} = \tau_{l-1}^2 + \alpha_{l-1,4}\chi_i + \alpha_{l-1,5}(t_i+\psi_i)^2 + \mathcal{O}\Big(p^{-3/2}\Big). \tag{23}$$

By Eq. (3), we have

$$\boldsymbol{G}_{ii}^{(l)} = \sigma_a^2 \mathbb{E}\left[\phi\left(\sqrt{\boldsymbol{G}_{ii}^{(l-1)}}\cdot\xi\right)^2\right] + (1-\sigma_a^2)\|\boldsymbol{x}_i\|^2,$$

for $\xi \sim \mathcal{N}(0,1)$.

By Taylor-expanding, one gets

$$\sqrt{\boldsymbol{G}_{ii}^{(l-1)}} = \tau_{l-1} + \frac{1}{2\tau_{l-1}}\alpha_{l-1,4}\chi_i + \frac{4\tau_{l-1}^2\alpha_{l-1,5} - \alpha_{l-1,4}^2}{8\tau_{l-1}^3}(t_i+\psi_i)^2 + \mathcal{O}\Big(p^{-3/2}\Big).$$

For simplicity, we denote the shortcut $f(\cdot) = \phi^2(\cdot)$. By Talor-expanding and Eq. (22), one gets

$$\begin{aligned}
\boldsymbol{G}_{ii}^{(l)} =& \sigma_a^2 \mathbb{E}\left[\phi\left(\sqrt{\boldsymbol{G}_{ii}^{(l-1)}}\cdot\xi\right)^2\right] + (1-\sigma_a^2)\|\boldsymbol{x}_i\|^2 = \sigma_a^2 \mathbb{E}\left[f\left(\sqrt{\boldsymbol{G}_{ii}^{(l-1)}}\cdot\xi\right)\right] + (1-\sigma_a^2)\|\boldsymbol{x}_i\|^2 \\
=& \sigma_a^2\mathbb{E}\left[f(\tau_{l-1}\xi) + f'(\tau_{l-1}\xi)\xi\left(\frac{1}{2\tau_{l-1}}\alpha_{l-1,4}\chi_i + \frac{4\tau_{l-1}^2\alpha_{l-1,5}-\alpha_{l-1,4}^2}{8\tau_{l-1}^3}(t_i+\psi_i)^2\right)\right] \\
&+ \sigma_a^2\mathbb{E}\left[\frac{1}{2}f''(\tau_{l-1}\xi)\xi^2\right]\frac{\alpha_{l-1,4}^2}{4\tau_{l-1}}(t_i+\psi_i)^2 + (1-\sigma_a^2)(\tau_0^2+\chi_i) + \mathcal{O}\Big(p^{-3/2}\Big) \\
=& \sigma_a^2\mathbb{E}\left[f(\tau_{l-1}\xi)\right] + (1-\sigma_a^2)\tau_0^2 + \left(\sigma_a^2\frac{\alpha_{l-1,4}}{2}\mathbb{E}[f''(\tau_{l-1}\xi)] + 1 - \sigma_a^2\right)\chi_i \\
&+ \sigma_a^2\frac{4\alpha_{l-1,5}\mathbb{E}[f''(\tau_{l-1}\xi)] + \alpha_{l-1,4}^2\mathbb{E}[f''''(\tau_{l-1}\xi)]}{8}(t_i+\psi_i)^2 + \mathcal{O}\Big(p^{-3/2}\Big),
\end{aligned}$$

where we use the facts that

$$\mathbb{E}[f''(\tau_{l-1}\xi)] = \tau_{l-1}\mathbb{E}[f''(\tau_{l-1}\xi)], \quad \mathbb{E}[f''''(\tau_{l-1}\xi)(\xi^2-1)] = \tau_{l-1}^2\mathbb{E}[f''''(\tau_{l-1}\xi)],$$

for $\xi \sim \mathcal{N}(0,1)$, as a result of the Gaussian integration by parts formula.

Consequently, we obtain the following relation

$$\begin{aligned}
\tau_l^2 &= \sigma_a^2\mathbb{E}\left[f(\tau_{l-1}\xi)\right] + (1-\sigma_a^2)\tau_0^2, \\
\alpha_{l,4} &= \frac{\sigma_a^2}{2}\mathbb{E}[f''(\tau_{l-1}\xi)]\alpha_{l-1,4} + 1 - \sigma_a^2, \\
\alpha_{l,5} &= \frac{\sigma_a^2}{2}\mathbb{E}[f''(\tau_{l-1}\xi)]\alpha_{l-1,5} + \frac{\sigma_a^2}{8}\mathbb{E}[f''''(\tau_{l-1}\xi)]\alpha_{l-1,4}^2.
\end{aligned} \tag{24}$$

By Lemma A.1, under Condition 2, it holds that $\frac{\sigma_a^2}{2}\mathbb{E}[f''(\tau_{l-1}\xi)] < 1$, which implies that, as $l\to\infty$, the iterations in Eq. (24) converge. Let $l\to\infty$, we obtain that

$$\begin{aligned}
\tau_* &\equiv \lim_{l\to\infty}\tau_l = \sqrt{\sigma_a^2\mathbb{E}\left[f(\tau_*\xi)\right] + (1-\sigma_a^2)\tau_0^2}, \\
\alpha_{*,4} &\equiv \lim_{l\to\infty}\alpha_{l,4} = \left(1 - \frac{\sigma_a^2}{2}\mathbb{E}[f''(\tau_*\xi)]\right)^{-1}(1-\sigma_a^2) \\
\alpha_{*,5} &\equiv \lim_{l\to\infty}\alpha_{l,5} = \frac{\sigma_a^2}{8}\left(1 - \frac{\sigma_a^2}{2}\mathbb{E}[f''(\tau_*\xi)]\right)^{-1}\mathbb{E}[f''''(\tau^*\xi)]\alpha_{*,4}^2.
\end{aligned} \tag{25}$$

Off the diagonal. For $i \neq j$, by induction hypothesis on the layer $l-1$, we have

$$\boldsymbol{G}_{ij}^{(l-1)} = \alpha_{l-1,1} A_{ij} + \alpha_{l-1,2}(t_i+\psi_i)(t_j+\psi_j) + \alpha_{l-1,3}\left(\frac{1}{p}\boldsymbol{z}_i^\top \boldsymbol{z}_j\right)^2 + S_{ij} + \mathcal{O}\Big(p^{-3/2}\Big).$$

Using the Gram-Schmidt orthogonalization for standard Gaussian random variable, we write

$$\boldsymbol{G}_{ii}^{(l)} = \sigma_a^2 \mathbb{E}\left[\phi^2\left(\sqrt{\boldsymbol{G}_{ii}^{(l-1)}}\cdot \xi_i\right)\right] + (1-\sigma_a^2)\|\boldsymbol{x}_i\|^2,$$

$$\boldsymbol{G}_{ij}^{(l)} = \sigma_a^2 \mathbb{E}\left[\phi\left(\sqrt{\boldsymbol{G}_{ii}^{(l-1)}}\cdot\xi_i\right)\times\phi\left(\frac{\boldsymbol{G}_{ij}^{(l-1)}}{\sqrt{\boldsymbol{G}_{ii}^{(l-1)}}}\cdot\xi_i + \sqrt{\boldsymbol{G}_{jj}^{(l-1)} - \frac{\left(\boldsymbol{G}_{ij}^{(l-1)}\right)^2}{\boldsymbol{G}_{ii}^{(l-1)}}}\cdot\xi_j\right)\right] + (1-\sigma_a^2)\boldsymbol{x}_i^\top\boldsymbol{x}_j.$$

Using Lemma B.1, we have

$$\begin{aligned}
\boldsymbol{G}_{ij}^{(l)} =& \sigma_a^2 \mathbb{E}\left[\phi\left(\sqrt{\boldsymbol{G}_{ii}^{(l-1)}}\cdot\xi_i\right)\times\phi\left(\frac{\boldsymbol{G}_{ij}^{(l-1)}}{\sqrt{\boldsymbol{G}_{ii}^{(l-1)}}}\cdot\xi_i + \sqrt{\boldsymbol{G}_{jj}^{(l-1)} - \frac{\left(\boldsymbol{G}_{ij}^{(l-1)}\right)^2}{\boldsymbol{G}_{ii}^{(l-1)}}}\cdot\xi_j\right)\right] + (1-\sigma_a^2)\boldsymbol{x}_i^\top\boldsymbol{x}_j \\
=& \sigma_a^2\mathbb{E}[\phi'(\tau_{l-1}\xi)]^2\boldsymbol{G}_{ij}^{(l-1)} \\
&+ \sigma_a^2\left(\frac{\alpha_{l-1,1}}{2}\mathbb{E}[\phi''(\tau_{l-1}\xi)]^2\left(\frac{1}{p}\boldsymbol{z}_i^\top\boldsymbol{z}_j\right)^2 + \frac{\alpha_{l-1,4}^2}{4}\mathbb{E}[\phi''(\tau_{l-1}\xi)]^2(t_i+\psi_i)(t_j+\psi_j)\right) \\
&+ S_{ij} + (1-\sigma_a^2)\boldsymbol{x}_i^\top\boldsymbol{x}_j + \mathcal{O}\Big(p^{-3/2}\Big) \\
=& \sigma_a^2\mathbb{E}[\phi'(\tau_{l-1}\xi)]^2\left(\alpha_{l-1,1}\boldsymbol{x}_i^\top\boldsymbol{x}_j + \alpha_{l-1,2}(t_i+\psi_i)(t_j+\psi_j) + \alpha_{l-1,3}\left(\frac{1}{p}\boldsymbol{z}_i^\top\boldsymbol{z}_j\right)^2\right) \\
&+ \sigma_a^2\left(\frac{\alpha_{l-1,1}^2}{2}\mathbb{E}[\phi''(\tau_{l-1}\xi)]^2\left(\frac{1}{p}\boldsymbol{z}_i^\top\boldsymbol{z}_j\right)^2 + \frac{\alpha_{l-1,4}^2}{4}\mathbb{E}[\phi''(\tau_{l-1}\xi)]^2(t_i+\psi_i)(t_j+\psi_j)\right) \\
&+ S_{ij} + (1-\sigma_a^2)\boldsymbol{x}_i^\top\boldsymbol{x}_j + \mathcal{O}\Big(p^{-3/2}\Big).
\end{aligned}$$

Consequently, it holds that

$$\boldsymbol{G}_{ij}^{(l)} = \alpha_{l,1}\boldsymbol{x}_i^\top\boldsymbol{x}_j + \alpha_{l,2}(t_i+\psi_i)(t_j+\psi_j) + \alpha_{l,3}\left(\frac{1}{p}\boldsymbol{z}_i^\top\boldsymbol{z}_j\right)^2 + S_{ij} + \mathcal{O}\Big(p^{-3/2}\Big), \tag{26}$$

where

$$\begin{aligned}
\alpha_{l,1} &= \sigma_a^2\mathbb{E}[\phi'(\tau_{l-1}\xi)]^2\alpha_{l-1,1} + 1 - \sigma_a^2, \\
\alpha_{l,2} &= \sigma_a^2\mathbb{E}[\phi'(\tau_{l-1}\xi)]^2\alpha_{l-1,2} + \frac{\sigma_a^2}{4}\mathbb{E}[\phi''(\tau_{l-1}\xi)]^2\alpha_{l-1,4}^2, \\
\alpha_{l,3} &= \sigma_a^2\mathbb{E}[\phi'(\tau_{l-1}\xi)]^2\alpha_{l-1,3} + \frac{\sigma_a^2}{2}\mathbb{E}[\phi''(\tau_{l-1}\xi)]^2\alpha_{l-1,1}^2.
\end{aligned} \tag{27}$$

Note that it holds that $\sigma_a^2\mathbb{E}[\phi'(\tau_*\xi)]^2 < 1$ under Condition 1. This means that, as $l \to \infty$, the iterations in Eq. (27) converge.Let $l \to \infty$, we obtain that

$$\begin{aligned}
\alpha_{*,1} &= \lim_{l\to\infty}\alpha_{l,1} = (1-\sigma_a^2\mathbb{E}[\phi'(\tau_*\xi)]^2)^{-1}(1-\sigma_a^2), \\
\alpha_{*,2} &= \lim_{l\to\infty}\alpha_{l,2} = \frac{\sigma_a^2}{4}(1-\sigma_a^2\mathbb{E}[\phi'(\tau_*\xi)]^2)^{-1}\mathbb{E}[\phi''(\tau_*\xi)]^2\alpha_{*,4}^2, \\
\alpha_{*,3} &= \lim_{l\to\infty}\alpha_{l,3} = \frac{\sigma_a^2}{2}(1-\sigma_a^2\mathbb{E}[\phi'(\tau_*\xi)]^2)^{-1}\mathbb{E}[\phi''(\tau_*\xi)]^2\alpha_{*,1}^2
\end{aligned} \tag{28}$$

where

$$\begin{aligned}
\tau_* &= \sqrt{\sigma_a^2\mathbb{E}\left[\phi^2(\tau_*\xi)\right] + (1-\sigma_a^2)\tau_0^2}, \\
\alpha_{*,4} &= \left(1 - \frac{\sigma_a^2}{2}\mathbb{E}[f''(\tau_*\xi)]\right)^{-1}(1-\sigma_a^2).
\end{aligned}$$

Assembling in matrix form. Using the fact that $\|M\|_2 \leq n \max_{i,j} |M_{ij}|$ for $M \in \mathbb{R}^{n\times n}$ and $\{S_{ij}\}_{ij} = \mathcal{O}_{\|\cdot\|}(p^{-1/2})$ (Couillet and Benaych-Georges, 2016), it holds that

$$G^* = \alpha_{*,1} X^\top X + VCV^\top + (\tau_*^2 - \tau_0^2 \alpha_{*,1} - \tau_0^4 \alpha_{*,3}) I_n + \mathcal{O}_{\|\cdot\|}(p^{-\frac{1}{2}}) \tag{29}$$

where $\mathcal{O}_{\|\cdot\|}(p^{-\frac{1}{2}})$ denotes matrices of spectral norm order $\mathcal{O}\Big(p^{-\frac{1}{2}}\Big)$, with

$$V = [J/\sqrt{p}, \psi], \quad C = \begin{bmatrix} \alpha_{*,2} tt^\top + \alpha_{*,3} T & \alpha_{*,2} t \\ \alpha_{*,2} t^\top & \alpha_{*,2} \end{bmatrix}, \tag{30}$$

and

$$T = \left\{ \frac{1}{p} \operatorname{tr} C_a C_b \right\}_{a,=1}^{K}, \quad t = \left\{ \frac{1}{\sqrt{p}} \operatorname{tr} C_a^\circ \right\}. \tag{31}$$

Remark 4 (Lack of bias term in Eq. (1) and Theorems 1–2)**.** Note that different from (Jacot et al., 2018; Feng and Kolter, 2020), here we consider (explicit or implicit) networks *without* the bias term. see Eq. (1). This is indeed a limitation of the present analysis approach, and the proposed theoretical framework is *not* able to cover deterministic and/or random bias. As a matter of fact, considering (say deterministic) bias in Eq. (1) is equivalent, from a technical perspective, to relax the "centered" activation assumption $\mathbb{E}[\phi(\tau_* \xi)] = 0, \xi \sim \mathcal{N}(0,1)$. This will make a few terms of order $\mathcal{O}\big(p^{-3/2}\big)$ no longer neglectable in the current proof of Theorem 1. To the best of our knowledge, the only work on precise high-dimensional asymptotics of DNN models that has taken the bias into account is (Adlam et al., 2022), but only on a single-hidden-layer and explicit neural network model. It would be of future interest to extend the proposed analysis approach to cover deterministic or random bias, which may lead to further improvement on the network practical performance.

C Proof of Theorem 2

C.1 The CK $\dot{G}$

Before proving Theorem 2, one needs to deal with the CK $\dot{G}$.

Recall that

$$\dot{G}_{ij}^{(l)} = \sigma_a^2 \mathbb{E}_{(\mathrm{u},\mathrm{v}) \sim \mathcal{N}(0, \Lambda_{ij}^{(l)})} [\phi'(\mathrm{u}) \phi'(\mathrm{v})],$$

where $\Lambda_{ij}^{(l)} = \begin{bmatrix} G_{ii}^{(l-1)} & G_{ij}^{(l-1)} \\ G_{ji}^{(l-1)} & G_{jj}^{(l-1)} \end{bmatrix}$.

Using the Gram-Schmidt orthogoalizaiton procedure, we have

$$\begin{aligned} \dot{G}_{ii}^{(l)} &= \sigma_a^2 \mathbb{E}\left[\phi'\left(\sqrt{G_{ii}^{(l-1)}} \cdot \xi_i \right)^2 \right] \\ \dot{G}_{ij}^{(l)} &= \sigma_a^2 \mathbb{E}\left[\phi'\left(\sqrt{G_{ii}^{(l-1)}} \cdot \xi_i \right) \times \phi'\left(\frac{G_{ij}^{(l-1)}}{\sqrt{G_{ii}^{(l-1)}}} \cdot \xi_i + \sqrt{G_{jj}^{(l-1)} - \frac{\left(G_{ij}^{(l-1)}\right)^2}{G_{ii}^{(l-1)}}} \cdot \xi_j \right) \right] \end{aligned} \tag{32}$$

On the diagonal. First recall that

$$\sqrt{G_{ii}^{(l-1)}} = \tau_{l-1} + \frac{1}{2\tau_{l-1}} \alpha_{l-1,4} \chi_i + \frac{4\tau_{l-1}^2 \alpha_{l-1,5} - \alpha_{l-1,4}^2}{8\tau_{l-1}^3} (t_i + \psi_i)^2 + \mathcal{O}\Big(p^{-3/2}\Big).$$

Denote the shortcut $f(t) = (\phi'(t))^2$, using Taylor-expand again, we have

$$
\begin{aligned}
\dot{\boldsymbol{G}}_{ii}^{(l)} =& \sigma_a^2 \mathbb{E}\left[\phi'\left(\sqrt{\boldsymbol{G}_{ii}^{(l-1)}} \cdot \xi\right)^2\right] = \sigma_a^2 \mathbb{E}\left[f\left(\sqrt{\boldsymbol{G}_{ii}^{(l-1)}} \cdot \xi\right)\right] \\
=& \sigma_a^2 \mathbb{E}\left[f(\tau_{l-1}\xi) + f'(\tau_{l-1}\xi)\xi\left(\frac{1}{2\tau_{l-1}}\alpha_{l-1,4}\chi_i + \frac{4\tau_{l-1}^2\alpha_{l-1,5} - \alpha_{l-1,4}^2}{8\tau_{l-1}^3}(t_i + \psi_i)^2\right)\right] \\
&+ \sigma_a^2 \mathbb{E}\left[\frac{1}{2} f''(\tau_{l-1}\xi)\xi^2\right] \frac{\alpha_{l-1,4}^2}{4\tau_{l-1}^2}(t_i + \psi_i)^2 + \mathcal{O}\left(p^{-3/2}\right) \\
=& \sigma_a^2 \mathbb{E}\left[f(\tau_{l-1}\xi)\right] + \left(\sigma_a^2 \frac{\alpha_{l-1,4}}{2} \mathbb{E}[f''(\tau_{l-1}\xi)]\right)\chi_i \\
&+ \sigma_a^2 \frac{4\alpha_{l-1,5}\mathbb{E}[f''(\tau_{l-1}\xi)] + \alpha_{l-1,4}^2 \mathbb{E}[f''''(\tau_{l-1}\xi)]}{8} + \mathcal{O}\left(p^{-3/2}\right),
\end{aligned}
$$

Thus, we conclude that

$$
\dot{\boldsymbol{G}}_{ii}^{(l)} = \sigma_a^2 \mathbb{E}\left[\phi'\left(\sqrt{\boldsymbol{G}_{ii}^{(l-1)}}\xi\right)^2\right] = \sigma_a^2 \dot{\tau}_l^2 + \mathcal{O}\left(p^{-1/2}\right),
$$

with the sequence $\dot{\tau}_l$ defined as follows

$$
\dot{\tau}_l = \sqrt{\mathbb{E}\left[\phi'(\tau_l\xi)^2\right]}.
$$

Off the diagonal For $i \neq j$, by Lemma B.1, it holds that

$$
\begin{aligned}
&\mathbb{E}\left[\phi'\left(\sqrt{\boldsymbol{G}_{ii}^{(l-1)}} \cdot \xi_i\right) \times \phi'\left(\frac{\boldsymbol{G}_{ij}^{(l-1)}}{\sqrt{\boldsymbol{G}_{ii}^{(l-1)}}} \cdot \xi_i + \sqrt{\boldsymbol{G}_{jj}^{(l-1)} - \frac{\left(\boldsymbol{G}_{ij}^{(l-1)}\right)^2}{\boldsymbol{G}_{ii}^{(l-1)}}} \cdot \xi_j\right)\right] \\
=& \mathbb{E}[\phi'(\tau_{l-1}\xi)]^2 + \mathbb{E}[\phi''(\tau_{l-1}\xi)]^2 \cdot \alpha_{l-1,1}\boldsymbol{x}_i^\top \boldsymbol{x}_j \\
&+ E[\phi'(\tau_{l-1}\xi)]E[\phi'''(\tau_{l-1}\xi)] \cdot \frac{\alpha_{l-1,4}}{2}(\chi_i + \chi_j) + \mathcal{O}\left(p^{-1}\right)
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
\dot{\boldsymbol{G}}_{ij}^{(l)} &= \sigma_a^2 \mathbb{E}\left[\phi'\left(\sqrt{\boldsymbol{G}_{ii}^{(l-1)}} \cdot \xi_i\right) \times \phi'\left(\frac{\boldsymbol{G}_{ij}^{(l-1)}}{\sqrt{\boldsymbol{G}_{ii}^{(l-1)}}} \cdot \xi_i + \sqrt{\boldsymbol{G}_{jj}^{(l-1)} - \frac{\left(\boldsymbol{G}_{ij}^{(l-1)}\right)^2}{\boldsymbol{G}_{ii}^{(l-1)}}} \cdot \xi_j\right)\right] \\
&= \sigma_a^2 \mathbb{E}[\phi'(\tau_{l-1}\xi)]^2 + \sigma_a^2 \mathbb{E}[\phi''(\tau_{l-1}\xi)]^2 \cdot \alpha_{l-1,1}\boldsymbol{x}_i^\top \boldsymbol{x}_j \\
&+ \sigma_a^2 E[\phi'(\tau_{l-1}\xi)]E[\phi'''(\tau_{l-1}\xi)] \cdot \frac{\alpha_{l-1,4}}{2}(\chi_i + \chi_j) + \mathcal{O}\left(p^{-1}\right) \\
&= \dot{\alpha}_{l,0} + \dot{\alpha}_{l,0}\boldsymbol{x}_i^\top \boldsymbol{x}_j + \dot{\alpha}_{l,0}(\chi_i + \chi_j) + \mathcal{O}\left(p^{-1}\right),
\end{aligned}
$$

with

$$
\begin{aligned}
\dot{\alpha}_{l,0} &= \sigma_a^2 \mathbb{E}[\phi'(\tau_{l-1}\xi)]^2, \\
\dot{\alpha}_{l,1} &= \sigma_a^2 \mathbb{E}[\phi''(\tau_{l-1}\xi)]^2 \alpha_{l-1,1}, \\
\dot{\alpha}_{l,2} &= \frac{\sigma_a^2}{2} \mathbb{E}[\phi'(\tau_{l-1}\xi)]\mathbb{E}[\phi'''(\tau_{l-1}\xi)]\alpha_{l-1,4}.
\end{aligned}
$$

As $l \to \infty$, $\lim_{l\to\infty} \tau_l = \tau_*$, and thus it holds that

$$
\begin{aligned}
\dot{\alpha}_{*,0} &= \sigma_a^2 \mathbb{E}[\phi'(\tau_*\xi)]^2, \\
\dot{\alpha}_{*,1} &= \sigma_a^2 \mathbb{E}[\phi''(\tau_*\xi)]^2 \alpha_{*,1}, \\
\dot{\alpha}_{*,2} &= \frac{\sigma_a^2}{2} \mathbb{E}[\phi'(\tau_*\xi)]\mathbb{E}[\phi'''(\tau_*\xi)]\alpha_{*,4}.
\end{aligned}
$$

C.2 IMPLICIT NTKS

Now, we are ready to prove Theorem 2. We assume the induction hypothesis holds for $l-1$, that

$$\boldsymbol{K}_{ii}^{(l)} = \kappa_{l-1} + \mathcal{O}\Big(p^{-1/2}\Big),$$
$$\boldsymbol{K}_{ij}^{(l)} = \beta_{l-1,1}\boldsymbol{x}_i^\top \boldsymbol{x}_j + \beta_{l-1,2}(t_i+\psi_i)(t_i+\psi_i) + \beta_{l-1,3}\left(\frac{1}{p}\boldsymbol{z}_i^\top \boldsymbol{z}_j\right)^2 + S_{ij} + \mathcal{O}\Big(p^{-3/2}\Big).$$

On the diagonal. First, note that

$$\boldsymbol{K}_{ii}^{(l)} = \boldsymbol{G}_{ii}^{(l)} + \boldsymbol{K}_{ii}^{(l-1)} \cdot \dot{\boldsymbol{G}}_{ii}^{(l)} = \tau_l^2 + \sigma_a^2 \kappa_l^2 \dot{\tau}_l^2 + \mathcal{O}\Big(p^{-1/2}\Big).$$

Thus, it holds that

$$\kappa_l^2 = \tau_l^2 + \sigma_a^2 \dot{\tau}_l^2 \cdot \kappa_{l-1}^2.$$

Under Condition 1, it holds that $\sigma_a^2 \mathbb{E}\left[\phi'(\tau_*\xi)^2\right] < 1$. Thus, for $l \to \infty$, one gets that

$$\kappa_*^2 = \left(1 - \sigma_a^2 \mathbb{E}\left[\phi'(\tau_*\xi)^2\right]\right)^{-1} \tau_*^2. \tag{33}$$

Off the diagonal. For $i \neq j$, we have

$$\begin{aligned}
\boldsymbol{K}_{ij}^{(l)} =& \boldsymbol{G}_{ij}^{(l)} + \boldsymbol{K}_{ij}^{(l-1)} \dot{\boldsymbol{G}}_{ij}^{(l)} \\
=& \alpha_{l,1}\boldsymbol{x}_i^\top \boldsymbol{x}_j + \alpha_{l,2}(t_i+\psi_i)(t_j+\psi_j) + \alpha_{l,3}\left(\frac{1}{p}\boldsymbol{z}_i^\top \boldsymbol{z}_j\right)^2 \\
&+ \left(\beta_{l-1,1}\boldsymbol{x}_i^\top \boldsymbol{x}_j + \beta_{l-1,2}(t_i+\psi_i)(t_j+\psi_j) + \beta_{l-1,3}\left(\frac{1}{p}\boldsymbol{z}_i^\top \boldsymbol{z}_j\right)^2\right) \\
&\times \left(\dot{\alpha}_{l,0} + \dot{\alpha}_{l,0}\boldsymbol{x}_i^\top \boldsymbol{x}_j + \dot{\alpha}_{l,0}\left(\chi_i + \chi_j\right) + \mathcal{O}(p^{-1})\right) + \mathcal{O}\Big(p^{-3/2}\Big) \\
=& (\alpha_{l,1} + \beta_{l-1,1} \cdot \dot{\alpha}_{l,0})\boldsymbol{x}_i^\top \boldsymbol{x}_j + (\alpha_{l,2} + \beta_{l-1,2} \cdot \dot{\alpha}_{l,0})(t_i+\psi_i)(t_j+\psi_j) \\
&+ (\alpha_{l,3} + \beta_{l-1,3} \cdot \dot{\alpha}_{l,0} + \beta_{l-1,1} \cdot \dot{\alpha}_{l,1})\left(\frac{1}{p}\boldsymbol{z}_i^\top \boldsymbol{z}_j\right)^2 + S_{ij} + \mathcal{O}\Big(p^{-3/2}\Big),
\end{aligned}$$

so that

$$\begin{aligned}
\beta_{l,1} &= \alpha_{l,1} + \beta_{l-1,1}\dot{\alpha}_{l,0}, \\
\beta_{l,2} &= \alpha_{l,2} + \beta_{l-1,2}\dot{\alpha}_{l,0}, \\
\beta_{l,3} &= \alpha_{l,3} + \beta_{l-1,3}\dot{\alpha}_{l,0} + \beta_{l-1,1}\dot{\alpha}_{l,1},
\end{aligned}$$

with

$$\begin{aligned}
\dot{\alpha}_{l,0} &= \sigma_a^2 \mathbb{E}[\phi'(\tau_{l-1}\xi)]^2 \\
\dot{\alpha}_{l,1} &= \sigma_a^2 \mathbb{E}[\phi''(\tau_{l-1}\xi)]^2 \alpha_{l-1,1}
\end{aligned}$$

As $l \to \infty$, it holds that $\lim_{l\to\infty} \tau_l = \tau_*$, $\lim_{l\to\infty} \alpha_{l,k} = \alpha_{*,k}$, and $\lim_{l\to\infty} \dot{\alpha}_{l,k} = \dot{\alpha}_{*,k}$, for $k = 1,2,3$. Therefore, for $l \to \infty$, one gets that

$$\boldsymbol{K}_{ij}^{(*)} = \beta_{*,1}\boldsymbol{x}_i^\top \boldsymbol{x}_j + \beta_{*,2}(t_i+\psi_i)(t_j+\psi_j) + \beta_{*,3}\left(\frac{1}{p}\boldsymbol{z}_i^\top \boldsymbol{z}_j\right)^2 + S_{ij} + \mathcal{O}\Big(p^{-3/2}\Big)$$

with

$$\begin{aligned}
\beta_{*,1} &= (1-\dot{\alpha}_{*,0})^{-1}\alpha_{*,1}, \\
\beta_{*,2} &= (1-\dot{\alpha}_{*,0})^{-1}\alpha_{*,2}, \\
\beta_{*,3} &= (1-\dot{\alpha}_{*,0})^{-1}(\alpha_{*,3} + \beta_{*,1}\dot{\alpha}_{*,1}),
\end{aligned}$$

and

$$\begin{aligned}
\dot{\alpha}_{*,0} &= \sigma_a^2 \mathbb{E}[\phi'(\tau_*\xi)]^2, \\
\dot{\alpha}_{*,1} &= \sigma_a^2 \mathbb{E}[\phi''(\tau_*\xi)]^2 \alpha_{*,1}.
\end{aligned}$$

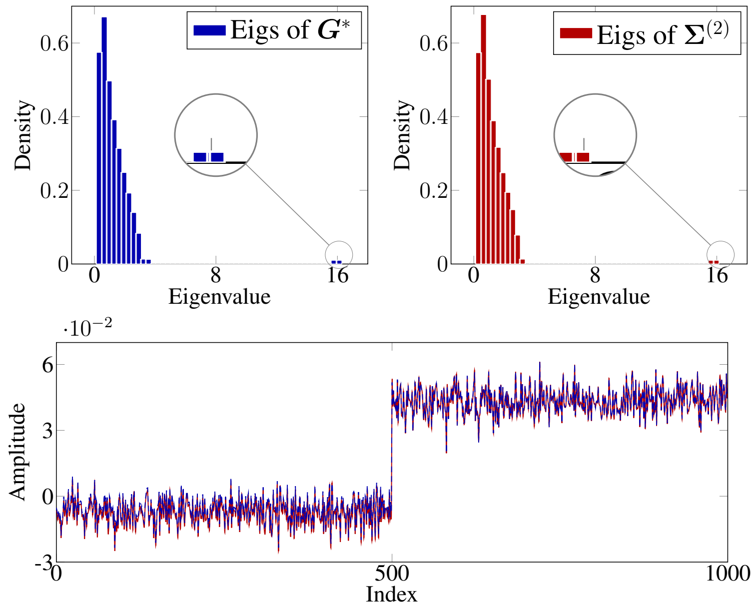

Figure 4: Eigenvalue density distribution histograms (**top**) and dominant eigenvectors (**bottom**) of Implicit-CK matrices $\boldsymbol{G}^*$ (**blue**) defined in Eq. (3) (with expectation estimated from 400 independent realizations of $\boldsymbol{A}$s and $\boldsymbol{B}$s) and Explicit-CKs $\boldsymbol{\Sigma}^{(2)}$ (**red**). A Gaussian implicit NN on two-class GMM data, with $p = 1000$, $n = 800$, $\boldsymbol{\mu}_a = [\mathbf{0}_{8(a-1)}; 8; \mathbf{0}_{p-8a+7}]$, $\boldsymbol{C}_a = (1 + 8(a-1)/\sqrt{p})\boldsymbol{I}_p$, for $a \in \{1, 2\}$ using Cosine activations.

Assembling in matrix form: Using the fact that $\|\boldsymbol{M}\|_2 \leq n \max_{i,j} |\boldsymbol{M}_{ij}|$ for $\boldsymbol{M} \in \mathbb{R}^{n\times n}$ and $\{S_{ij}\}_{ij} = \mathcal{O}_{\|\cdot\|}(p^{-1/2})$ Couillet and Benaych-Georges (2016), it holds that

$$\boldsymbol{K}^* = \beta_{*,1}\boldsymbol{X}^\top\boldsymbol{X} + \boldsymbol{V}\boldsymbol{D}_*\boldsymbol{V}^\top + (\kappa_*^2 - \tau_0^2\beta_{*,1} - \tau_0^4\beta_{*,3})\boldsymbol{I}_n + \mathcal{O}_{\|\cdot\|}(p^{-\frac{1}{2}}), \tag{34}$$

with

$$\boldsymbol{V} = [\boldsymbol{J}/\sqrt{p}, \boldsymbol{\psi}], \quad \boldsymbol{D}_* = \begin{bmatrix} \beta_{*,2}\boldsymbol{t}\boldsymbol{t}^\top + \beta_{*,3}\boldsymbol{T} & \beta_{*,2}\boldsymbol{t} \\ \beta_{*,2}\boldsymbol{t}^\top & \beta_{*,2} \end{bmatrix}, \tag{35}$$

and

$$\boldsymbol{T} = \left\{\frac{1}{p}\operatorname{tr}\boldsymbol{C}_a\boldsymbol{C}_b\right\}_{a,=1}^K, \quad \boldsymbol{t} = \left\{\frac{1}{\sqrt{p}}\operatorname{tr}\boldsymbol{C}_a^\circ\right\}. \tag{36}$$

D Proof of Corollary 1

It follows from (Gu et al., 2022, Theorem 1) that for weight matrices $\boldsymbol{W}_l$s having *i.i.d.* entries of zero mean, unit variance, and finite fourth-order moment, one has, for any two-hidden-layer fully-connected *explicit* NN defined in Eq. (18) that, the second-layer CK matrix $\boldsymbol{\Sigma}^{(2)}$ satisfies

$$\boldsymbol{\Sigma}^{(2)} = \tilde{\alpha}_{2,1}\boldsymbol{X}^\top\boldsymbol{X} + \boldsymbol{V}\tilde{\boldsymbol{C}}_2\boldsymbol{V}^\top + (\tilde{\tau}_2^2 - \tilde{\tau}_0^2\tilde{\alpha}_{2,1} - \tilde{\tau}_0^4\tilde{\alpha}_{2,3})\boldsymbol{I}_n + \mathcal{O}_{\|\cdot\|}(n^{-1/2})$$

where $\tilde{\tau}_2^2 = \mathbb{E}[\sigma_1^2(\tilde{\tau}_1\xi)]$, $\tilde{\tau}_1^2 = \mathbb{E}[\sigma_1^2(\tau_0\xi)]$ for $\xi \sim \mathcal{N}(0,1)$, and $\tilde{\tau}_0 = \tau_0$, $\sigma_l(\cdot)$ are "centred" such that $\mathbb{E}[\sigma_l(\tilde{\tau}_l\xi)] = 0$ for $l = 1, 2$, and

$$\boldsymbol{V} = [\boldsymbol{J}/\sqrt{p},\ \boldsymbol{\psi}] \in \mathbb{R}^{n\times(K+1)}, \quad \tilde{\boldsymbol{C}}_2 = \begin{bmatrix} \tilde{\alpha}_{2,2}\boldsymbol{t}\boldsymbol{t}^\top + \tilde{\alpha}_{2,3}\boldsymbol{T} & \tilde{\alpha}_{2,2}\boldsymbol{t} \\ \tilde{\alpha}_{2,2}\boldsymbol{t}^\top & \tilde{\alpha}_{2,2} \end{bmatrix} \in \mathbb{R}^{(K+1)\times(K+1)} \tag{37}$$

with

$$\tilde{\alpha}_{2,1} = \mathbb{E}[\sigma_2'(\tilde{\tau}_1\xi)]^2\mathbb{E}[\sigma_1'(\tilde{\tau}_0\xi)]^2,\ \tilde{\alpha}_{2,2} = \frac{1}{4}\mathbb{E}[\sigma_2'(\tilde{\tau}_1\xi)]^2\mathbb{E}[\sigma_1''(\tilde{\tau}_0\xi)]^2 + \frac{1}{8}\mathbb{E}[\sigma_2''(\tilde{\tau}_1\xi)]^2\mathbb{E}[(\sigma_1^2(\tau_0\xi))'']^2,$$

$$\tilde{\alpha}_{2,3} = \frac{1}{2}\mathbb{E}[\sigma_2'(\tilde{\tau}_1\xi)]^2\mathbb{E}[\sigma_1''(\tilde{\tau}_0\xi)]^2 + \frac{1}{2}\mathbb{E}[\sigma_2''(\tilde{\tau}_1\xi)]^2\mathbb{E}[\sigma_1'(\tilde{\tau}_0\xi)]^4.$$

m	32	64	128	256	512	1024	2048	4096	8192
$\\|\boldsymbol{G}^* - \boldsymbol{\Sigma}^{(2)}\\|$	17.32	10.24	6.53	3.71	1.60	0.93	0.81	0.12	0.12

Table 1: The spectral norm difference $\|\boldsymbol{G}^* - \boldsymbol{\Sigma}^{(2)}\|$ for finite-width ploy-ENN on GMM data under the same setting as Figure 4.

Since the Explicit-CK $\boldsymbol{\Sigma}^{(2)}$ takes a similar form as the "equivalent" Implicit-CK $\overline{\boldsymbol{G}}$ (and thus the Implicit-CK $\boldsymbol{G}^*$) per Theorem 1, it then suffices to choose the activations of ploy-ENN in such a way that the key coefficients $\widetilde{\alpha}_1, \widetilde{\alpha}_2, \widetilde{\alpha}_3$ and $\widetilde{\tau}_2^2$ coincide with those $\alpha_{*,1}, \widetilde{\alpha}_{*,2}, \widetilde{\alpha}_{*,3}$ and τ_*^2 defined in Theorem 1. Specifically, We consider a quadratic function $\sigma_l(x) = a_l x^2 + b_l x + c_l$. In order to ensure $\sigma_l(\cdot)$ are "centered," *i.e.,* $\mathbb{E}[\sigma_l(\widetilde{\tau}_{l-1}\xi)] = 0$, we take $c_l = -a_l\widetilde{\tau}_{l-1}^2$ such that

$$\sigma_l(x) = a_l x^2 + b_l x - a_l \widetilde{\tau}_{l-1}^2.$$

One can derive that, for $\xi \sim \mathcal{N}(0,1)$,

$$\mathbb{E}[\sigma_l^2(\widetilde{\tau}_{l-1}\xi)] = 2a_l^2\widetilde{\tau}_{l-1}^4 + b_l^2\widetilde{\tau}_{l-1}^2, \quad \mathbb{E}[(\sigma_l^2(\widetilde{\tau}_{l-1}))''] = 8a_l^2\widetilde{\tau}_{l-1}^2 + 2b_l^2,$$
$$\mathbb{E}[\sigma_l'(\widetilde{\tau}_{l-1}\xi)]^2 = b_l^2, \quad \mathbb{E}[\sigma_l''(\widetilde{\tau}_{l-1}\xi)]^2 = 4a_l^2.$$

As a result, we have

$$\widetilde{\alpha}_{2,1} = b_2^2 b_1^2, \quad \widetilde{\alpha}_{2,2} = b_2^2 a_1^2 + a_2^2(4a_1^2\tau_0^2 + b_1^2)^2,$$
$$\widetilde{\alpha}_{2,3} = 2a_1^2 b_2^2 + 2a_2^2 b_1^2, \quad \widetilde{\tau}_2^2 = 2a_2^2(2a_1^2\tau_0^4 + b_1^2\tau_0^2)^2 + b_2^2(2a_1^2\tau_0^4 + b_1^2\tau_0^2).$$

Solving a_1, b_1, a_2, b_2, such that $\widetilde{\alpha}_{2,1} = \alpha_{*,1}, \widetilde{\alpha}_{2,2} = \alpha_{*,2}, \widetilde{\alpha}_{2,3} = \alpha_{*,3}$, and $\tau_*^2 - \tau_0^2\alpha_{*,1} - \tau_0^4\alpha_{*,3} = \widetilde{\tau}_2^2 - \widetilde{\tau}_0^2\widetilde{\alpha}_1 - \widetilde{\tau}_0^4\widetilde{\alpha}_3$, we obtain the aimed activations σ_1 and σ_2 of two-layer "equivalent" explicit NNs. This concludes the proof of Corollary 1.

A numerical simulation using Cosine activations is presented in Figure 4 to validate our theory. Moreover, Table 1 demonstrates that $\|\boldsymbol{G}^* - \boldsymbol{\Sigma}^{(2)}\|$ for finite-width ploy-ENN decreases with the increase of the width m. Note particularly that the approximation saturates at a low level (~ 0.12) and this is due to the finite n, p in the setting of Figure 4 as opposed to our asymptotic theoretical results.

E Proof of Corollary 2

For $\varphi(x) \equiv \max(ax, bx) - \frac{a-b}{\sqrt{2\pi}}\tau_0$, with $a \geq b \geq 0$, we have, for $\xi \sim \mathcal{N}(0,1)$

$$\mathbb{E}[\varphi^2(\tau_0\xi)] = \frac{(a^2+b^2)(\pi-1) + 2ab}{2\pi}\tau_0^2, \quad \mathbb{E}[\varphi'(\tau_0\xi)]^2 = \frac{(a-b)^2}{2}.$$

The corresponding CK of the single-layer explicit NN with the activation φ is given by

$$\widetilde{G}_\varphi = \widetilde{\alpha}_1 \boldsymbol{X}^\top + (\widetilde{\tau}_2^2 - \widetilde{\tau}_1^2\widetilde{\alpha}_1 - \widetilde{\tau}_1^4\widetilde{\alpha}_3)\boldsymbol{I}_n,$$

where

$$\widetilde{\alpha}_1 = \mathbb{E}[\varphi'(\tau_0\xi)]^2, \quad \widetilde{\tau}_1^2 = \sqrt{\mathbb{E}[\varphi^2(\tau_0\xi)]}.$$

Solving a and b such that $\widetilde{\alpha}_1 = \alpha_{*,1}$ and $\widetilde{\tau}_1^2 - \tau_0^2\widetilde{\alpha}_1 = \tau_*^2 - \tau_0\alpha_{*,1}$, *i.e.,*

$$\begin{cases} (a-b)^2 = 4\alpha_{*,1}^2 \\ \frac{(\pi-1)(a^2+b^2)+ab}{2\pi}\tau_0^2 - \frac{(a+b)^2}{4}\tau_0^2 = \tau_*^2 - \alpha_{*,1}\tau_0^2 \end{cases},$$

we obtain the aimed biased Leaky-ReLU function φ of the single-layer equivalent explicit NN. This concludes the proof of Corollary 2.

F Additional Experimental Results

Here we provide additional experiments on the CIFAR-10 dataset. We use features from a pretrained VGG-19 model. The VGG-19 model is specifically pre-trained on the training set of CIFAR-10 data. This differs from the case of Figure 3-(c) which uses a VGG-19 pre-trained on ImageNet. As shown in Figure 5, the best accuracy is close to 85% (which is significantly higher that Figure 3-(c)), and as the dimension m increases, the performance of L-ReLU-ENNs closely matches that of INN. Meanwhile, a noticeable performance gap exists between ReLU-ENN and INN.

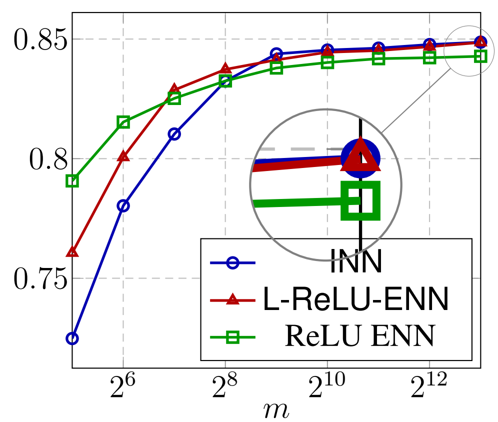

Figure 5: The evolution of classification results *w.r.t* the width m of implicit ReLU NNs (**blue**, INN), the corresponding equivalent single-layer Leaky-ReLU explicit NNs (**red**, L-ReLU-ENN), and ReLU explicit NNs (**green**, ReLU ENN for short) on CIFAR-10. Different from Figure 3-(c) which uses a VGG-19 pre-trained on ImageNet, here, we adopt a VGG-19 pretrained on the training set of CIFAR-10, and get (significantly) higher accuracy than Figure 3-(c),