A Large width asymptotics: $k = 1$

We first consider the case with of a single input being a real-valued vector of dimension I. By means of (4) and (5) we can write for $i \geq 1$ and $l = 2, \ldots, D$:

i)

$$\begin{aligned}\varphi_{f_i^{(1)}(x)}(t) &= \mathbb{E}[e^{\mathrm{i}tf_i^{(1)}(x)}] \\ &= \mathbb{E}\left[\exp\left\{\mathrm{i}t\left[\sum_{j=1}^{I} w_{i,j}^{(1)}x_j + b_i^{(1)}\right]\right\}\right] \\ &= \mathbb{E}[\exp\{\mathrm{i}tb_i^{(1)}\}]\prod_{j=1}^{I}\mathbb{E}[\exp\{\mathrm{i}tw_{i,j}^{(1)}x_j\}] \\ &= \mathrm{e}^{-\sigma_b^\alpha|t|^\alpha}\prod_{j=1}^{I}\mathrm{e}^{-\sigma_w^\alpha|tx_j|^\alpha} \\ &= \exp\left\{-(\sigma_w^\alpha\sum_{j=1}^{I}|x_j|^\alpha + \sigma_b^\alpha)|t|^\alpha\right\},\end{aligned}$$

i.e.,

$$f_i^{(1)}(x) \stackrel{\mathrm{d}}{=} S_{\alpha,\left(\sigma_w^\alpha\sum_{j=1}^{I}|x_j|^\alpha+\sigma_b^\alpha\right)^{1/\alpha}};$$

ii)

$$\begin{aligned}&\varphi_{f_i^{(l)}(x,n)\,|\,\{f_j^{(l-1)}(x,n)\}_{j=1,\ldots,n}}(t) \\ &= \mathbb{E}[e^{\mathrm{i}tf_i^{(l)}(x,n)}\,|\,\{f_j^{(l-1)}(x,n)\}_{j=1,\ldots,n}] \\ &= \mathbb{E}\left[\exp\left\{\mathrm{i}t\left[\frac{1}{n^{1/\alpha}}\sum_{j=1}^{n} w_{i,j}^{(l)}\phi(f_j^{(l-1)}(x,n)) + b_i^{(l)}\right]\right\}\,|\,\{f_j^{(l-1)}(x,n)\}_{j=1,\ldots,n}\right] \\ &= \mathbb{E}[\exp\{\mathrm{i}tb_i^{(l)}\}]\prod_{j=1}^{n}\mathbb{E}[\exp\{\mathrm{i}tw_{i,j}^{(l)}\frac{\phi(f_j^{(l-1)}(x,n))}{n^{1/\alpha}}\,|\,\{f_j^{(l-1)}(x,n)\}_{j=1,\ldots,n}\}] \\ &= \mathrm{e}^{-\sigma_b^\alpha|t|^\alpha}\prod_{j=1}^{n}\mathrm{e}^{-\frac{\sigma_w^\alpha}{n}|t\phi(f_j^{(l-1)}(x,n))|^\alpha} \\ &= \exp\left\{-(\frac{\sigma_w^\alpha}{n}\sum_{j=1}^{n}|\phi(f_j^{(l-1)}(x,n))|^\alpha + \sigma_b^\alpha)|t|^\alpha\right\},\end{aligned}$$

i.e.,

$$f_i^{(l)}(x,n)\,|\,\{f_j^{(l-1)}(x,n)\}_{j=1,\ldots,n} \stackrel{\mathrm{d}}{=} S_{\alpha,\left(\frac{\sigma_w^\alpha}{n}\sum_{j=1}^{n}|\phi(f_j^{(l-1)}(x,n))|^\alpha+\sigma_b^\alpha\right)^{1/\alpha}}. \tag{18}$$

We show that, as $n \to +\infty$,

$$f_i^{(l)}(x,n) \xrightarrow{w} \mathrm{St}(\alpha, \sigma(l)), \tag{19}$$

and we determine the expression of $\sigma(l)$.

A.1 Asymptotics for the i-th coordinate

It comes from (8) that, for every fixed l and for every fixed n the sequence $(f_i^{(l)}(n,x))_{i\geq 1}$ is exchangeable. In particular, let $p_n^{(l)}$ denote the directing (random) probability measure of the exchangeable sequence $(f_i^{(l)}(n,x))_{i\geq 1}$.

That is, by de Finetti representation theorem, conditionally to $p_n^{(l)}$ the $f_i^{(l)}(n,x)$'s are iid as $p_n^{(l)}$. Now, consider the induction hypothesis that, as $n \to +\infty$

$$p_n^{(l-1)} \xrightarrow{w} q^{(l-1)},$$

with $q^{(l-1)}$ being $\mathrm{St}(\alpha, \sigma(l-1))$, and the parameter $\sigma(l-1)$ will be specified. Therefore, we can write the following expression

$$\begin{aligned}
\mathbb{E}[\mathrm{e}^{\mathrm{i}tf_i^{(l)}(x,n)}] &= \mathbb{E}\left[\exp\left\{-|t|^\alpha \left(\frac{\sigma_w^\alpha}{n}\sum_{j=1}^n |\phi(f_j^{(l-1)}(x,n))|^\alpha + \sigma_b^\alpha\right)\right\}\right] \\
&= \exp\{-|t|^\alpha \sigma_b^\alpha\}\,\mathbb{E}\left[\exp\left\{-|t|^\alpha \frac{\sigma_w^\alpha}{n}\sum_{j=1}^n |\phi(f_j^{(l-1)}(x,n))|^\alpha\right\}\right] \\
&= \exp\{-|t|^\alpha \sigma_b^\alpha\}\,\mathbb{E}\left[\left(\int \exp\left\{-|t|^\alpha \frac{\sigma_w^\alpha}{n}|\phi(f)|^\alpha\right\} p_n^{(l-1)}(\mathrm{d}f)\right)^n\right].
\end{aligned} \tag{20}$$

Hereafter we show the limiting behaviour (19). In order to prove this limiting behaviour, we will prove:

L1) for each $l \geq 2$ $\Pr[p_n^{(l-1)} \in I] = 1$, with $I = \{p : \int |\phi(f)|^\alpha p(\mathrm{d}f) < +\infty\}$;

L1.1) for each $l \geq 2$ there exists $\epsilon > 0$ such that $\sup_n \mathbb{E}[|\phi(f_i^{(l-1)}(x,n))|^{\alpha+\epsilon} \mid p_n^{(l-2)}] < +\infty$;

L2) $\int |\phi(f)|^\alpha p_n^{(l-1)}(\mathrm{d}f) \xrightarrow{p} \int |\phi(f)|^\alpha q^{(l-1)}(\mathrm{d}f)$, as $n \to +\infty$;

L3) $\int |\phi(f)|^\alpha [1 - e^{-|t|^\alpha \frac{\sigma_w^\alpha}{n}|\phi(f)|^\alpha}] p_n^{(l-1)}(\mathrm{d}f) \xrightarrow{p} 0$, as $n \to +\infty$.

A.1.1 Proof of L1)

The proof of L1) follows by induction. In particular, L1) is true for the envelope condition (6), i.e.,

$$\begin{aligned}
\mathbb{E}[|\phi(f_i^{(1)}(x))|^\alpha] &\leq \mathbb{E}[(a + b|f_i^{(1)}(x)|^\beta)^{\gamma\alpha}] \\
&\leq \mathbb{E}[a^{\gamma\alpha} + b^{\gamma\alpha}|f_i^{(1)}(x)|^{\beta\gamma\alpha}] \\
&= a^{\gamma\alpha} + b^{\gamma\alpha}\mathbb{E}[|f_i^{(1)}(x)|^{\beta\gamma\alpha}] \\
&< +\infty,
\end{aligned}$$

since $\alpha\beta\gamma < \alpha$. Now, assuming that L1) is true for $(l-2)$, we prove that it is true for $(l-1)$. Again from (6), one has

$$\begin{aligned}
&\mathbb{E}[|\phi(f_i^{(l-1)}(x,n))|^\alpha \mid \{f_j^{(l-2)}(x,n)\}_{j=1,\dots,n}] \\
&\quad \leq \mathbb{E}[(a + b|f_i^{(l-1)}(x,n)|^\beta)^{\gamma\alpha} \mid \{f_j^{(l-2)}(x,n)\}_{j=1,\dots,n}] \\
&\quad \leq \mathbb{E}[a^{\gamma\alpha} + b^{\gamma\alpha}\ |f_i^{(l-1)}(x,n)|^{\beta\gamma\alpha} \mid \{f_j^{(l-2)}(x,n)\}_{j=1,\dots,n}] \\
&\quad = a^{\gamma\alpha} + b^{\gamma\alpha}\mathbb{E}[|f_i^{(l-1)}(x,n)|^{\beta\gamma\alpha} \mid \{f_j^{(l-2)}(x,n)\}_{j=1,\dots,n}] \\
&\quad \leq a^{\gamma\alpha} + b^{\gamma\alpha}\mathbb{E}[|S_{\alpha,1}|^{\alpha\beta\gamma}] \left(\frac{\sigma_w^\alpha}{n}\sum_{j=1}^n |\phi(f_j^{(l-2)}(x,n))|^\alpha + \sigma_b^\alpha\right)^{\beta\gamma}.
\end{aligned}$$

Thus, since $\beta < \gamma^{-1}$,

$$\begin{aligned}
&\mathbb{E}[|\phi(f_i^{(l-1)}(x,n))|^\alpha \mid p^{(l-2)}(x,n)] \\
&\quad = \mathbb{E}[\mathbb{E}[|\phi(f_i^{(l-1)}(x,n))|^\alpha \mid \{f_j^{(l-2)}(x,n)\}_{j=1,\dots,n}] \mid p_n^{(l-2)}] \\
&\quad \leq a^{\gamma\alpha} + b^{\gamma\alpha}\mathbb{E}[|S_{\alpha,1}|^{\alpha\beta\gamma}]\mathbb{E}\left[\left(\frac{\sigma_w^\alpha}{n}\sum_{j=1}^n |\phi(f_j^{(l-2)}(x,n))|^\alpha + \sigma_b^\alpha\right)^{\beta\gamma} \mid p_n^{(l-2)}\right]
\end{aligned}$$

$$
\begin{aligned}
&\leq a^{\gamma\alpha} + b^{\gamma\alpha}\mathbb{E}[|S_{\alpha,1}|^{\alpha\beta\gamma}] \left(\mathbb{E}\left[\frac{\sigma_w^\alpha}{n}\sum_{j=1}^{n}|\phi(f_j^{(l-2)}(x,n))|^\alpha + \sigma_b^\alpha \mid p_n^{(l-2)}\right]\right)^{\beta\gamma} \\
&\leq a^{\gamma\alpha} + b^{\gamma\alpha}\mathbb{E}[|S_{\alpha,1}|^{\alpha\beta\gamma}] \left(\sigma_b^\alpha + \sigma_w^\alpha \int |\phi(f)|^\alpha p_n^{(l-2)}(\mathrm{d}f)\right)^{\beta\gamma} \\
&< +\infty.
\end{aligned}
$$

A.1.2 Proof of L1.1)

The proof of L1.1) follows by induction, and along lines similar to the proof of L1). In particular, let ϵ be such that $\beta\gamma(\alpha+\epsilon)/\alpha < 1$ and $\gamma(\alpha+\epsilon) < 1$. It exists since $\beta\gamma < 1$ and $\gamma\alpha < 1$. For $l = 2$,

$$
\begin{aligned}
\mathbb{E}(|\phi(f_i^{(1)}(x))|^{\alpha+\epsilon}) &\leq \mathbb{E}[(a + b|f_i^{(1)}(x)|^\beta)^{\gamma(\alpha+\epsilon)}] \\
&\leq \mathbb{E}[a^{\gamma(\alpha+\epsilon)} + b^{\gamma(\alpha+\epsilon)}|f_i^{(1)}(x)|^{(\alpha+\epsilon)\gamma\beta}] \\
&= a^{\gamma(\alpha+\epsilon)} + b^{\gamma(\alpha+\epsilon)}\mathbb{E}[|f_i^{(1)}(x)|^{(\alpha+\epsilon)\gamma\beta}] \\
&< +\infty,
\end{aligned}
$$

since $(\alpha+\epsilon)\beta\gamma < \alpha$. This follows along lines similar to those applied in the previous subsection. Moreover the bound is uniform with respect to n since the law is invariant with respect to n. Now, assume that L1.1) is true for $(l-2)$. Then, we can write the following inequality

$$
\begin{aligned}
&\mathbb{E}[|\phi(f_i^{(l-1)}(x,n))|^{\alpha+\epsilon} \mid p_n^{(l-2)}] \\
&\quad\leq a^{\gamma(\alpha+\epsilon)} + b^{\gamma(\alpha+\epsilon)}\mathbb{E}[|S_{\alpha,1}|^{\beta\gamma(\alpha+\epsilon)}] \left(\mathbb{E}\left[\frac{\sigma_w^\alpha}{n}\sum_{j=1}^{n}|\phi(f_j^{(l-2)}(x,n))|^\alpha + \sigma_b^\alpha \mid p_n^{(l-2)}\right]\right)^{\beta\gamma(\alpha+\epsilon)/\alpha} \\
&\quad\leq a^{\gamma(\alpha+\epsilon)} + b^{\gamma(\alpha+\epsilon)}\mathbb{E}[|S_{\alpha,1}|^{\beta\gamma(\alpha+\epsilon)}] \left(\sigma_b^\alpha + \sigma_w^\alpha \int |\phi(f)|^\alpha p_n^{(l-2)}(\mathrm{d}f)\right)^{\beta\gamma(\alpha+\epsilon)/\alpha}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\sup_n \mathbb{E}[|\phi(f_i^{(l-1)}(x,n))|^{\alpha+\epsilon} \mid p_n^{(l-2)}] \\
&\quad\leq a^{\gamma(\alpha+\epsilon)} + b^{\gamma(\alpha+\epsilon)}\mathbb{E}[|S_{\alpha,1}|^{\beta\gamma(\alpha+\epsilon)}] \left(\sigma_b^\alpha + \sigma_w^\alpha \sup_n \left[\int |\phi(f)|^\alpha p_n^{(l-2)}(\mathrm{d}f)\right]\right)^{\beta\gamma(\alpha+\epsilon)/\alpha} \\
&\quad< +\infty.
\end{aligned}
$$

A.1.3 Proof of L2)

By the induction hypothesis, $p_n^{(l-1)}$ converges to $p^{(l-1)}$ in distribution with respect to the weak topology. Since the limit law is degenerate (on $p^{(l-1)}$), then for every subsequence (n') there exists a subsequence (n'') such that $p_{n''}^{(l-1)}$ converges a.s. By the induction hypothesis, $p^{(l-1)}$ is absolutely continuous with respect to the Lebesgue measure. Since $|\phi|^\alpha$ is a.s. continuous and, by L1.1), uniformly integrable with respect to $(p_n^{(l-1)})$, then we can write the following

$$
\int |\phi(f)|^\alpha p_{n''}^{(l-1)}(\mathrm{d}f) \longrightarrow \int |\phi(f)|^\alpha q^{(l-1)}(\mathrm{d}f) \quad a.s.
$$

Thus, $n \to +\infty$

$$
\int |\phi(f)|^\alpha p_n^{(l-1)}(\mathrm{d}f) \xrightarrow{p} \int |\phi(f)|^\alpha q^{(l-1)}(\mathrm{d}f).
$$

A.1.4 Proof of L3)

Let ϵ be as in L1.1), and let $p = (\alpha+\epsilon)/\alpha$ and $q = (\alpha+\epsilon)/\epsilon$. Then $1/p + 1/q = 1$. Thus, by Holder inequality

$$
\int |\phi(f)|^\alpha [1 - e^{-|t|^\alpha \frac{\sigma_w^\alpha}{n}|\phi(f)|^\alpha}] p_n^{(l-1)}(\mathrm{d}f)
$$

$$\leq \left(\int |\phi(f)|^{\alpha p} p_n^{(l-1)}(\mathrm{d}f)\right)^{1/p} \left(\int [1-e^{-|t|^\alpha \frac{\sigma_w^\alpha}{n}|\phi(f)|^\alpha}]^q p_n^{(l-1)}(\mathrm{d}f)\right)^{1/q}.$$

Since we defined $p = (\alpha+\epsilon)/\alpha$ and $q = (\alpha+\epsilon)/\epsilon$, i.e. we set $q > 1$, then we can write the following

$$\begin{aligned}
&\left(\int |\phi(f)|^{\alpha p} p_n^{(l-1)}(\mathrm{d}f)\right)^{1/p} \left(\int [1-e^{-|t|^\alpha \frac{\sigma_w^\alpha}{n}|\phi(f)|^\alpha}]^q p_n^{(l-1)}(\mathrm{d}f)\right)^{1/q} \\
&\quad \leq \sup_n \left[\left(\int |\phi(f)|^{\alpha+\epsilon} p_n^{(l-1)}(\mathrm{d}f)\right)^{1/p}\right] \left(\int [1-e^{-|t|^\alpha \frac{\sigma_w^\alpha}{n}|\phi(f)|^\alpha}] p_n^{(l-1)}(\mathrm{d}f)\right)^{1/q} \\
&\quad \leq \sup_n \left[\left(\int |\phi(f)|^{\alpha+\epsilon} p_n^{(l-1)}(\mathrm{d}f)\right)^{1/p}\right] \left(|t|^\alpha \frac{\sigma_w^\alpha}{n} \int |\phi(f)|^\alpha p_n^{(l-1)}(\mathrm{d}f)\right)^{1/q} \longrightarrow 0,
\end{aligned}$$

as $n \to \infty$, by L1.1).

A.1.5 Combine L1), L2) and L3)

We combine L1), L2) and L3) to prove the large n behavior of the i-th coordinate $n^{-1/\alpha} f_i(x, n)$. From (20)

$$\begin{aligned}
&\mathbb{E}[\mathrm{e}^{\mathrm{i}tf_i^{(l)}(x,n)}] \\
&\quad = \exp\{-|t|^\alpha \sigma_b^\alpha\} \mathbb{E}\left[\left(\int \exp\left\{-|t|^\alpha \frac{\sigma_w^\alpha}{n}|\phi(f)|^\alpha\right\} p_n^{(l-1)}(\mathrm{d}f)\right)^n\right] \\
&\quad = \exp\{-|t|^\alpha \sigma_b^\alpha\} \mathbb{E}\left[\mathbb{1}_{\{(p_n^{(l-1)} \in I)\}}\left(\int \exp\left\{-|t|^\alpha \frac{\sigma_w^\alpha}{n}|\phi(f)|^\alpha\right\} p_n^{(l-1)}(\mathrm{d}f)\right)^n\right].
\end{aligned}$$

Then, by Lagrange theorem, there exists a value $\theta_n \in [0, 1]$ such that the following equality holds true

$$1 - \exp\left\{-|t|^\alpha \frac{\sigma_w^\alpha}{n}|\phi(f)|^\alpha\right\} = |t|^\alpha \frac{\sigma_w^\alpha}{n}|\phi(f)|^\alpha \exp\left\{-\theta_n |t|^\alpha \frac{\sigma_w^\alpha}{n}|\phi(f)|^\alpha\right\},$$

thus

$$\begin{aligned}
&\exp\left\{-|t|^\alpha \frac{\sigma_w^\alpha}{n}|\phi(f)|^\alpha\right\} \\
&\quad = 1 - |t|^\alpha \frac{\sigma_w^\alpha}{n}|\phi(f)|^\alpha \exp\left\{-\theta_n |t|^\alpha \frac{\sigma_w^\alpha}{n}|\phi(f)|^\alpha\right\} \\
&\quad = 1 - |t|^\alpha \frac{\sigma_w^\alpha}{n}|\phi(f)|^\alpha + |t|^\alpha \frac{\sigma_w^\alpha}{n}|\phi(f)|^\alpha \left(1 - \exp\left\{-\theta_n |t|^\alpha \frac{\sigma_w^\alpha}{n}|\phi(f)|^\alpha\right\}\right).
\end{aligned}$$

Now, since

$$\begin{aligned}
0 &\leq \int |\phi(f)|^\alpha [1 - e^{-\theta_n |t|^\alpha \frac{\sigma_w^\alpha}{n}|\phi(f)|^\alpha}] p_n^{(l-1)}(\mathrm{d}f) \\
&\leq \int |\phi(f)|^\alpha [1 - e^{-|t|^\alpha \frac{\sigma_w^\alpha}{n}|\phi(f)|^\alpha}] p_n^{(l-1)}(\mathrm{d}f),
\end{aligned}$$

then

$$\begin{aligned}
&\mathbb{E}[\mathrm{e}^{\mathrm{i}tf_i^{(l)}(x,n)}] \\
&\quad \leq \mathbb{E}[\exp\{-|t|^\alpha \sigma_b^\alpha\}]\mathbb{E}\Bigg[\mathbb{1}_{\{(p_n^{(l-1)} \in I)\}}\Bigg(1 - |t|^\alpha \frac{\sigma_w^\alpha}{n}\int |\phi(f)|^\alpha p_n^{(l-1)}(\mathrm{d}f) \\
&\qquad\qquad + |t|^\alpha \frac{\sigma_w^\alpha}{n}\int |\phi(f)|^\alpha [1 - e^{-|t|^\alpha \frac{\sigma_w^\alpha}{n}|\phi(f)|^\alpha}] p_n^{(l-1)}(\mathrm{d}f)\Bigg)^n\Bigg].
\end{aligned}$$

Thus, by using the definition of the exponential function, i.e. $\mathrm{e}^x = \lim_{n\to+\infty}(1 + x/n)^n$, and L1)-L3) we have

$$\mathbb{E}[\mathrm{e}^{\mathrm{i}tf_i^{(l)}(x,n)}] \to e^{-|t|^\alpha[\sigma_b^\alpha + \sigma_w^\alpha \int |\phi(f)|^\alpha q^{(l-1)}(\mathrm{d}f)]},$$

as $n \to +\infty$. That is, we proved that the large n limiting distribution of $f_i^{(l)}(x, n)$ is $\mathrm{St}(\alpha, \sigma(l))$, where

$$\sigma(l) = \left(\sigma_b^\alpha + \sigma_w^\alpha \int |\phi(f)|^\alpha q^{(l-1)}(\mathrm{d}f) \right)^{1/\alpha}.$$

B Large width asymptotics: $k \geq 1$

We now consider the case with of a k inputs, each one being a real-valued vector of dimension I. We represent this generic case with a $I \times k$ input matrix $\boldsymbol{X}$. Let $\mathbf{1}_r$ denote a vector of dimension $k \times 1$ with 1 in the r-the entry and 0 elsewhere, and $\mathbf{1}$ denote a vector of dimension $k \times 1$ of 1's. If $\boldsymbol{x}_j$ denotes the j-th row of the input matrix, then we can write

$$f_i^{(1)}(\boldsymbol{X}, n) = \sum_{j=1}^{I} w_{i,j}^{(1)} \boldsymbol{x}_j + b_i^{(1)} \mathbf{1},$$

and

$$f_i^{(l)}(\boldsymbol{X}, n) = \frac{1}{n^\alpha} \sum_{j=1}^{n} w_{i,j}^{(l)} (\phi \circ f_j^{(l-1)}(\boldsymbol{X}, n)) + b_i^{(l)} \mathbf{1}$$

for $l = 2, \ldots, D$, $i = 1, \ldots, n$ where we denote with $\circ$ element-wise application. Note that $f_i^{(l)}(\boldsymbol{X}, n)$ is a random vector of dimension $k \times 1$, and we denote the r-th component of this vector by $f_{i,r}^{(l)}(\boldsymbol{X}, n)$, namely $f_{i,r}^{(l)}(\boldsymbol{X}, n) = \mathbf{1}_r^T f_i^{(l)}(\boldsymbol{X}, n)$. Then, by means of (4) and (5), we can write for $i = 1 \geq 1$ and $l = 2, \ldots, D$:

i)

$$\begin{aligned}
\varphi_{f_i^{(1)}(\boldsymbol{X},n)}(\boldsymbol{t}) &= \mathbb{E}[e^{\mathrm{i}\boldsymbol{t}^T f_i^{(1)}(\boldsymbol{X},n)}] \\
&= \mathbb{E}\left[\exp\left\{\mathrm{i}\boldsymbol{t}^T \left[\sum_{j=1}^{I} w_{i,j}^{(1)} \boldsymbol{x}_j + b_i^{(1)} \mathbf{1}\right]\right\}\right] \\
&= \mathbb{E}[\exp\{\mathrm{i}\boldsymbol{t}^T b_i^{(1)} \mathbf{1}\}] \prod_{j=1}^{I} \mathbb{E}[\exp\{\mathrm{i}\boldsymbol{t}^T w_{i,j}^{(1)} \boldsymbol{x}_j\}] \\
&= \mathrm{e}^{-\sigma_b^\alpha |\boldsymbol{t}^T \mathbf{1}|^\alpha} \prod_{j=1}^{I} \mathrm{e}^{-\sigma_w^\alpha |\boldsymbol{t}^T \boldsymbol{x}_j|^\alpha} \\
&= \exp\left\{-\sigma_b^\alpha ||\mathbf{1}||^\alpha \left|\boldsymbol{t}^T \frac{\mathbf{1}}{||\mathbf{1}||}\right|^\alpha\right\} \exp\left\{-\sigma_w^\alpha \sum_{j=1}^{I} ||\boldsymbol{x}_j||^\alpha \left|\boldsymbol{t}^T \frac{\boldsymbol{x}_j}{||\boldsymbol{x}_j||}\right|^\alpha\right\} \\
&= \exp\left\{-\int_{\mathbb{S}^{k-1}} |\boldsymbol{t}^T \boldsymbol{s}|^\alpha \left(||\sigma_b \mathbf{1}||^\alpha \delta_{\frac{\mathbf{1}}{||\mathbf{1}||}} + \sum_{j=1}^{I} ||\sigma_w \boldsymbol{x}_j||^\alpha \delta_{\frac{\boldsymbol{x}_j}{||\boldsymbol{x}_j||}}\right)(\mathrm{d}\boldsymbol{s})\right\},
\end{aligned}$$

i.e.,

$$f_i^{(1)}(\boldsymbol{X}) \stackrel{\mathrm{d}}{=} S_{\alpha,\Gamma^{(1)}}$$

with

$$\Gamma^{(1)} = ||\sigma_b \mathbf{1}||^\alpha \delta_{\frac{\mathbf{1}}{||\mathbf{1}||}} + \sum_{j=1}^{I} ||\sigma_w \boldsymbol{x}_j||^\alpha \delta_{\frac{\boldsymbol{x}_j}{||\boldsymbol{x}_j||}};$$

observe that we can also determine the marginal distributions of $f_i^{(1)}(\boldsymbol{X})$. From (2), (3),

$$f_{i,r}^{(1)}(\boldsymbol{X}) \sim \mathrm{St}(\alpha, \sigma^{(1)}(r)),$$

with

$$\sigma^{(1)}(r) = \left(\int_{\mathbb{S}^{k-1}} |\mathbf{1}_r^T \boldsymbol{s}|^\alpha \Gamma^{(1)}(\mathrm{d}\boldsymbol{s})\right)^{1/\alpha}$$

ii)

$$\varphi_{f_i^{(l)}(\boldsymbol{X},n) \,|\, \{f_j^{(l-1)}(\boldsymbol{X},n)\}_{j=1,\ldots,n}}(\boldsymbol{t})$$

$$
\begin{aligned}
&= \mathbb{E}[e^{\mathrm{i}\boldsymbol{t}^T f_i^{(l)}(\boldsymbol{X},n)} \,|\, \{f_j^{(l-1)}(\boldsymbol{X},n)\}_{j=1,\dots,n}] \\
&= \mathbb{E}\left[\exp\left\{\mathrm{i}\boldsymbol{t}^T\left[\frac{1}{n^{1/\alpha}}\sum_{j=1}^{n} w_{i,j}^{(l)}(\phi\circ f_j^{(l-1)}(\boldsymbol{X},n)) + b_i^{(l)}\boldsymbol{1}\right]\right\} \,|\, \{f_j^{(l-1)}(\boldsymbol{X},n)\}_{j=1,\dots,n}\right] \\
&= \mathbb{E}[\exp\{\mathrm{i}\boldsymbol{t}^T b_i^{(l)}\boldsymbol{1}\}]\prod_{j=1}^{n}\mathbb{E}[\exp\{\mathrm{i}\boldsymbol{t}^T w_{i,j}^{(l)}\frac{\phi\circ f_j^{(l-1)}(\boldsymbol{X},n)}{n^{1/\alpha}} \,|\, \{f_j^{(l-1)}(\boldsymbol{X},n)\}_{j=1,\dots,n}\}] \\
&= \mathrm{e}^{-\sigma_b^\alpha|\boldsymbol{t}^T\boldsymbol{1}|^\alpha}\prod_{j=1}^{n}\mathrm{e}^{-\frac{\sigma_w^\alpha}{n}|\boldsymbol{t}^T(\phi\circ f_j^{(l-1)}(\boldsymbol{X},n))|^\alpha} \\
&= \exp\left\{-\sigma_b^\alpha||\boldsymbol{1}||^\alpha\left|\boldsymbol{t}^T\frac{\boldsymbol{1}}{||\boldsymbol{1}||}\right|^\alpha\right\} \\
&\quad\times\exp\left\{-\frac{\sigma_w^\alpha}{n}\sum_{j=1}^{n}||\phi\circ f_j^{(l-1)}(\boldsymbol{X},n)||^\alpha\left|\boldsymbol{t}^T\frac{\phi\circ f_j^{(l-1)}(\boldsymbol{X},n)}{||\phi\circ f_j^{(l-1)}(\boldsymbol{X},n)||}\right|^\alpha\right\} \\
&= \exp\left\{-\int_{\mathbb{S}^{k-1}}|\boldsymbol{t}^T\boldsymbol{s}|^\alpha\left(||\sigma_b\boldsymbol{1}||^\alpha\delta_{\frac{\boldsymbol{1}}{||\boldsymbol{1}||}} + \frac{1}{n}\sum_{j=1}^{n}||\sigma_w(\phi\circ f_j^{(l-1)}(\boldsymbol{X},n))||^\alpha\delta_{\frac{\phi\circ f_j^{(l-1)}(\boldsymbol{X},n)}{||\phi\circ f_j^{(l-1)}(\boldsymbol{X},n)||}}\right)(\mathrm{d}\boldsymbol{s})\right\},
\end{aligned}
$$

i.e.,

$$
f_i^{(l)}(\boldsymbol{X},n)\,|\,\{f_j^{(l-1)}(\boldsymbol{X},n)\}_{j=1,\dots,n} \overset{\mathrm{d}}{=} S_{\alpha,\Gamma^{(l)}}
$$

with

$$
\Gamma^{(l)} = ||\sigma_b\boldsymbol{1}||^\alpha\delta_{\frac{\boldsymbol{1}}{||\boldsymbol{1}||}} + \frac{1}{n}\sum_{j=1}^{n}||\sigma_w(\phi\circ f_j^{(l-1)}(\boldsymbol{X},n))||^\alpha\delta_{\frac{\phi\circ f_j^{(l-1)}(\boldsymbol{X},n)}{||\phi\circ f_j^{(l-1)}(\boldsymbol{X},n)||}};
$$

observe that we can also determine the marginal distributions of $f_i^{(l)}(\boldsymbol{X})$. From (2), (3),

$$
f_{i,r}^{(l)}(\boldsymbol{X},n)\,|\,\{f_j^{(l-1)}(\boldsymbol{X},n)\}_{j=1,\dots,n} \sim \mathrm{St}(\alpha,\sigma^{(l)}(r)),
$$

with

$$
\sigma^{(l)}(r) = \left(\int_{\mathbb{S}^{k-1}}|\boldsymbol{1}_r^T\boldsymbol{s}|^\alpha\Gamma^{(l)}(\mathrm{d}\boldsymbol{s})\right)^{1/\alpha}. \tag{21}
$$

We show that, as $n\to+\infty$,

$$
f_i^{(l)}(\boldsymbol{X},n) \xrightarrow{w} \mathrm{St}_k(\alpha,\Gamma(l)), \tag{22}
$$

and we determine the expression of $\Gamma(l)$.

B.1 Asymptotics for the i-th coordinate

Let $p_n^{(l)}$ denote the directing (random) measure of the exchangeable sequence $(f_i^{(l)}(n,\boldsymbol{X}))_{i\geq 1}$. Now, consider the induction hypothesis that, as $n\to+\infty$,

$$
p_n^{(l-1)} \xrightarrow{w} q^{(l-1)},
$$

with $q^{(l-1)}$ being $\mathrm{St}(\alpha,\sigma(l-1))$, and the finite measure $\Gamma(l-1)$ will be specified. Therefore, we can write the following expression

$$
\begin{aligned}
&\mathbb{E}[\mathrm{e}^{\mathrm{i}\boldsymbol{t}^T f_i^{(l)}(\boldsymbol{X},n)}] \qquad (23)\\
&\quad= \mathbb{E}\left[\exp\left\{-\int_{\mathbb{S}^{k-1}}|\boldsymbol{t}^T\boldsymbol{s}|^\alpha\tilde{\Gamma}^{(l)}(\mathrm{d}\boldsymbol{s})\right\}\right] \\
&\quad= \mathbb{E}\left[\exp\left\{-\int_{\mathbb{S}^{k-1}}|\boldsymbol{t}^T\boldsymbol{s}|^\alpha\left(||\sigma_b\boldsymbol{1}||^\alpha\delta_{\frac{\boldsymbol{1}}{||\boldsymbol{1}||}}\right)(\mathrm{d}\boldsymbol{s})\right\}\right]
\end{aligned}
$$

$$\times\,\mathbb{E}\left[\exp\left\{-\int_{\mathbb{S}^{k-1}}|\boldsymbol{t}^T\boldsymbol{s}|^\alpha\left(\frac{1}{n}\sum_{j=1}^{n}||\sigma_w(\phi\circ f_j^{(l-1)}(\boldsymbol{X},n))||^\alpha\delta_{\frac{\phi\circ f_j^{(l-1)}(\boldsymbol{X},n)}{||\phi\circ f_j^{(l-1)}(\boldsymbol{X},n)||}}\right)(\mathrm{d}\boldsymbol{s})\right\}\right]$$
$$=\mathbb{E}\left[\exp\left\{-\int_{\mathbb{S}^{k-1}}|\boldsymbol{t}^T\boldsymbol{s}|^\alpha\left(||\sigma_b\mathbf{1}||^\alpha\delta_{\frac{\mathbf{1}}{||\mathbf{1}||}}\right)(\mathrm{d}\boldsymbol{s})\right\}\right]$$
$$\times\,\mathbb{E}\left[\left(\int\exp\left\{-\int_{\mathbb{S}^{k-1}}|\boldsymbol{t}^T\boldsymbol{s}|^\alpha\left(\frac{1}{n}||\sigma_w(\phi\circ f)||^\alpha\delta_{\frac{\phi\circ f}{||\phi\circ f||}}\right)(\mathrm{d}\boldsymbol{s})\right\}p_n^{(l-1)}(\mathrm{d}f)\right)^n\right].$$

Hereafter we show the limiting behaviour (22). In order to prove this limiting behaviour, we will prove:

L1) for each $l\geq 2$ $\Pr[p_n^{(l-1)}\in I]=1$, with $I=\{p:\int||\phi\circ f||^\alpha p(\mathrm{d}f)<+\infty\}$;

L1.1) for each $l\geq 2$ there exists $\epsilon>0$ such that $\sup_n\mathbb{E}[||\phi\circ f_i^{(l-1)}(\boldsymbol{X},n)||^{\alpha+\epsilon}\mid p_n^{(l-2)}]<+\infty$;

L2) $\int||\phi\circ f||^\alpha p_n^{(l-1)}(\mathrm{d}f)\xrightarrow{p}\int||\phi\circ f||^\alpha q^{(l-1)}(\mathrm{d}f)$, as $n\to+\infty$;

L3) $\int||\phi\circ f||^\alpha\left[1-\exp\left\{-\int_{\mathbb{S}^{k-1}}|\boldsymbol{t}^T\boldsymbol{s}|^\alpha\left(\frac{1}{n}||\sigma_w(\phi\circ f)||^\alpha\delta_{\frac{\phi\circ f}{||\phi\circ f||}}\right)(\mathrm{d}\boldsymbol{s})\right\}\right]p_n^{(l-1)}(\mathrm{d}f)\xrightarrow{p}0$, as $n\to+\infty$.

B.1.1 Proof of L1)

The proof of L1) follows by induction. In particular, L1) is true for the envelope condition (6), i.e.,

$$\begin{aligned}\mathbb{E}[||\phi\circ f_i^{(1)}(\boldsymbol{X})||^\alpha]&\leq\mathbb{E}\left[\sum_{r=1}^{k}|\phi(f_{i,r}^{(1)}(\boldsymbol{X}))|^\alpha\right]\\&\leq\sum_{r=1}^{k}\mathbb{E}[(a+b|f_{i,r}^{(1)}(\boldsymbol{X})|^\beta)^{\gamma\alpha}]\\&\leq\sum_{r=1}^{k}\mathbb{E}[(a^{\gamma\alpha}+b^{\gamma\alpha}|f_{i,r}^{(1)}(\boldsymbol{X})|^{\beta\gamma\alpha})]\\&\leq ka^{\gamma\alpha}+b^{\gamma\alpha}\sum_{r=1}^{k}\mathbb{E}[|f_{i,r}^{(1)}(\boldsymbol{X})|^{\beta\gamma\alpha}]\\&<+\infty\end{aligned}$$

since $\alpha\beta\gamma<\alpha$. Now, assuming that L1) is true for $(l-2)$, we prove that it is true for $(l-1)$. Again from (6),

$$\begin{aligned}&\mathbb{E}[||\phi\circ f_i^{(l-1)}(\boldsymbol{X},n)||^\alpha\,|\,\{f_j^{(l-2)}(\boldsymbol{X},n)\}_{j=1,\dots,n}]\\&\quad\leq\mathbb{E}\left[\sum_{r=1}^{k}|\phi(f_{i,r}^{(l-1)}(\boldsymbol{X},n))|^\alpha\,|\,\{f_j^{(l-2)}(\boldsymbol{X},n)\}_{j=1,\dots,n}\right]\\&\quad\leq\sum_{r=1}^{k}\mathbb{E}[(a+b\,|f_{i,r}^{(l-1)}(\boldsymbol{X},n)|^\beta)^{\gamma\alpha}\,|\,\{f_j^{(l-2)}(\boldsymbol{X},n)\}_{j=1,\dots,n}]\\&\quad\leq\sum_{r=1}^{k}\mathbb{E}[(a^{\gamma\alpha}+b^{\gamma\alpha}\,|f_{i,r}^{(l-1)}(\boldsymbol{X},n)|^{\beta\gamma\alpha})\,|\,\{f_j^{(l-2)}(\boldsymbol{X},n)\}_{j=1,\dots,n}]\\&\quad=ka^{\gamma\alpha}+b^{\gamma\alpha}\sum_{r=1}^{k}\mathbb{E}[|f_{i,r}^{(l-1)}(\boldsymbol{X},n)|^{\beta\gamma\alpha}\,|\,\{f_j^{(l-2)}(\boldsymbol{X},n)\}_{j=1,\dots,n}]\\&\quad\leq ka^{\gamma\alpha}+b^{\gamma\alpha}\mathbb{E}[|S_{\alpha,1}|^{\alpha\beta\gamma}]\\&\qquad\times\sum_{r=1}^{k}\left(\int_{\mathbb{S}^{k-1}}|\mathbf{1}_r^T\boldsymbol{s}|^\alpha\left(||\sigma_b\mathbf{1}||^\alpha\delta_{\frac{\mathbf{1}}{||\mathbf{1}||}}\right)(\mathrm{d}\boldsymbol{s})\right.\end{aligned}$$

$$+\int_{\mathbb{S}^{k-1}} |\mathbf{1}_r^T \boldsymbol{s}|^\alpha \left(\frac{1}{n}\sum_{j=1}^n ||\sigma_w(\phi\circ f_j^{(l-2)}(\boldsymbol{X},n))||^\alpha \delta_{\frac{\phi\circ f_j^{(l-2)}(\boldsymbol{X},n)}{||\phi\circ f_j^{(l-2)}(\boldsymbol{X},n)||}}\right)(\mathrm{d}\boldsymbol{s})\Bigg)^{\beta\gamma}.$$

Since $\beta < \gamma^{-1}$,

$$\begin{aligned}
&\mathbb{E}[||\phi\circ f_i^{(l-1)}(\boldsymbol{X},n)||^\alpha \,|\, p_n^{(l-2)}]\\
&= \mathbb{E}[\mathbb{E}[||\phi\circ f_i^{(l-1)}(\boldsymbol{X},n)||^\alpha \,|\, \{f_j^{(l-2)}(\boldsymbol{X},n)\}_{j=1,\dots,n}] \,|\, p_n^{(l-2)}]\\
&\le k a^{\gamma\alpha} + b^{\gamma\alpha}\mathbb{E}[|S_{\alpha,1}|^{\alpha\beta\gamma}]\\
&\quad\times \sum_{r=1}^k \mathbb{E}\Bigg[\Bigg(\int_{\mathbb{S}^{k-1}} |\mathbf{1}_r^T\boldsymbol{s}|^\alpha \left(||\sigma_b\mathbf{1}||^\alpha \delta_{\frac{\mathbf{1}}{||\mathbf{1}||}}\right)(\mathrm{d}\boldsymbol{s})\\
&\qquad + \int_{\mathbb{S}^{k-1}} |\mathbf{1}_r^T \boldsymbol{s}|^\alpha \left(\frac{1}{n}\sum_{j=1}^n ||\sigma_w(\phi\circ f_j^{(l-2)}(\boldsymbol{X},n))||^\alpha \delta_{\frac{\phi\circ f_j^{(l-2)}(\boldsymbol{X},n)}{||\phi\circ f_j^{(l-2)}(\boldsymbol{X},n)||}}\right)(\mathrm{d}\boldsymbol{s})\Bigg)^{\beta\gamma} \,|\, p_n^{(l-2)}\Bigg]\\
&\le k a^{\gamma\alpha} + b^{\gamma\alpha}\mathbb{E}[|S_{\alpha,1}|^{\alpha\beta\gamma}]\\
&\quad\times \sum_{r=1}^k \Bigg(\mathbb{E}\Bigg[\int_{\mathbb{S}^{k-1}} |\mathbf{1}_r^T\boldsymbol{s}|^\alpha \left(||\sigma_b\mathbf{1}||^\alpha \delta_{\frac{\mathbf{1}}{||\mathbf{1}||}}\right)(\mathrm{d}\boldsymbol{s})\\
&\qquad + \int_{\mathbb{S}^{k-1}} |\mathbf{1}_r^T \boldsymbol{s}|^\alpha \left(\frac{1}{n}\sum_{j=1}^n ||\sigma_w(\phi\circ f_j^{(l-2)}(\boldsymbol{X},n))||^\alpha \delta_{\frac{\phi\circ f_j^{(l-2)}(\boldsymbol{X},n)}{||\phi\circ f_j^{(l-2)}(\boldsymbol{X},n)||}}\right)(\mathrm{d}\boldsymbol{s}) \,|\, p_n^{(l-2)}\Bigg]\Bigg)^{\beta\gamma}\\
&\le k a^{\gamma\alpha} + b^{\gamma\alpha}\mathbb{E}[|S_{\alpha,1}|^{\alpha\beta\gamma}]\\
&\quad\times \sum_{r=1}^k \Bigg(\int_{\mathbb{S}^{k-1}} |\mathbf{1}_r^T\boldsymbol{s}|^\alpha \left(||\sigma_b\mathbf{1}||^\alpha \delta_{\frac{\mathbf{1}}{||\mathbf{1}||}}\right)(\mathrm{d}\boldsymbol{s})\\
&\qquad + \int_{\mathbb{S}^{k-1}} |\mathbf{1}_r^T \boldsymbol{s}|^\alpha \left(\int ||\sigma_w(\phi\circ f)||^\alpha \delta_{\frac{\phi\circ f}{||\phi\circ f||}} p_n^{(l-2)}(\mathrm{d}f)\right)(\mathrm{d}\boldsymbol{s})\Bigg)^{\beta\gamma}\\
&< +\infty.
\end{aligned}$$

B.1.2 Proof of L1.1)

The proof of L1.1) follows by induction, and along lines similar to the proof of L1). In particular, let ϵ be such that $\beta\gamma(\alpha+\epsilon)/\alpha < 1$ and $\gamma(\alpha+\epsilon) < 1$. It exists since $\beta\gamma < 1$ and $\gamma\alpha < 1$. For $l = 2$, we can find $C(k) > 0$ finite such that:

$$\begin{aligned}
\mathbb{E}(||\phi\circ f_i^{(1)}(\boldsymbol{X})||^{\alpha+\epsilon}) &\le \mathbb{E}\left[C(k)\sum_{r=1}^k |\phi(f_{i,r}^{(1)}(\boldsymbol{X}))|^{\alpha+\epsilon}\right]\\
&\le C(k)\sum_{r=1}^k \mathbb{E}[(a + b|f_{i,r}^{(1)}(\boldsymbol{X})|^\beta)^{\gamma(\alpha+\epsilon)}]\\
&\le C(k)\sum_{r=1}^k \mathbb{E}[(a^{\gamma(\alpha+\epsilon)} + b^{\gamma(\alpha+\epsilon)}|f_{i,r}^{(1)}(\boldsymbol{X})|^{\beta\gamma(\alpha+\epsilon)})]\\
&= C(k)\left(k a^{\gamma(\alpha+\epsilon)} + b^{\gamma(\alpha+\epsilon)}\sum_{r=1}^k \mathbb{E}[|f_{i,r}^{(1)}(\boldsymbol{X})|^{(\alpha+\epsilon)\gamma\beta}]\right)\\
&< +\infty,
\end{aligned}$$

since $(\alpha+\epsilon)\beta\gamma < \alpha$. This follows along lines similar to those applied in the previous subsection. Moreover the bound is uniform with respect to n since the law is invariant with respect to n. Let us assume that L1.1) is true for $(l-2)$. Then we can write the following

$$\begin{aligned}
&\mathbb{E}[||\phi\circ f_i^{(l-1)}(\boldsymbol{X},n)||^{\alpha+\epsilon}\mid\{f_j^{(l-2)}(x,n)\}_{j=1,\dots,n}]\\
&\quad\le C(k)\Bigg\{ka^{\gamma(\alpha+\epsilon)}+b^{\gamma(\alpha+\epsilon)}\mathbb{E}[|S_{\alpha,1}|^{\beta\gamma(\alpha+\epsilon)}]\\
&\qquad\times\sum_{r=1}^{k}\Bigg(\mathbb{E}\Bigg[\int_{\mathbb{S}^{k-1}}|\mathbf{1}_r^T\boldsymbol{s}|^{\alpha}\left(||\sigma_b\mathbf{1}||^{\alpha}\delta_{\frac{\mathbf{1}}{||\mathbf{1}||}}\right)(\mathrm{d}\boldsymbol{s})\\
&\qquad\qquad+\int_{\mathbb{S}^{k-1}}|\mathbf{1}_r^T\boldsymbol{s}|^{\alpha}\left(\frac{1}{n}\sum_{j=1}^{n}||\sigma_w(\phi\circ f_j^{(l-2)}(\boldsymbol{X},n))||^{\alpha}\delta_{\frac{\phi\circ f_j^{(l-2)}(\boldsymbol{X},n)}{||\phi\circ f_j^{(l-2)}(\boldsymbol{X},n)||}}\right)(\mathrm{d}\boldsymbol{s})\mid p_n^{(l-2)}\Bigg]\Bigg)^{\beta\gamma(\alpha+\epsilon)/\alpha}\Bigg\}\\
&\quad\le C(k)\Bigg\{ka^{\gamma(\alpha+\epsilon)}+b^{\gamma(\alpha+\epsilon)}\mathbb{E}[|S_{\alpha,1}|^{\beta\gamma(\alpha+\epsilon)}]\\
&\qquad\times\sum_{r=1}^{k}\Bigg(\int_{\mathbb{S}^{k-1}}|\mathbf{1}_r^T\boldsymbol{s}|^{\alpha}\left(||\sigma_b\mathbf{1}||^{\alpha}\delta_{\frac{\mathbf{1}}{||\mathbf{1}||}}\right)(\mathrm{d}\boldsymbol{s})\\
&\qquad\qquad+\int_{\mathbb{S}^{k-1}}|\mathbf{1}_r^T\boldsymbol{s}|^{\alpha}\left(\int||\sigma_w(\phi\circ f)||^{\alpha}\delta_{\frac{\phi\circ f}{||\phi\circ f||}}p_n^{(l-2)}(\mathrm{d}f)\right)(\mathrm{d}\boldsymbol{s})\Bigg)^{\beta\gamma(\alpha+\epsilon)/\alpha}\Bigg\}
\end{aligned}$$

Thus,

$$\begin{aligned}
&\sup_n\mathbb{E}[||\phi\circ f_i^{(l-1)}(\boldsymbol{X},n)||^{\alpha+\epsilon}\mid p_n^{(l-2)}]\\
&\quad\le C(k)\Bigg\{ka^{\gamma(\alpha+\epsilon)}+b^{\gamma(\alpha+\epsilon)}\mathbb{E}[|S_{\alpha,1}|^{\beta\gamma(\alpha+\epsilon)}]\\
&\qquad\times\sup_n\sum_{r=1}^{k}\Bigg(\int_{\mathbb{S}^{k-1}}|\mathbf{1}_r^T\boldsymbol{s}|^{\alpha}\left(||\sigma_b\mathbf{1}||^{\alpha}\delta_{\frac{\mathbf{1}}{||\mathbf{1}||}}\right)(\mathrm{d}\boldsymbol{s})\\
&\qquad\qquad+\int_{\mathbb{S}^{k-1}}|\mathbf{1}_r^T\boldsymbol{s}|^{\alpha}\left(\int||\sigma_w(\phi\circ f)||^{\alpha}\delta_{\frac{\phi\circ f}{||\phi\circ f||}}p_n^{(l-2)}(\mathrm{d}f)\right)(\mathrm{d}\boldsymbol{s})\Bigg)^{\beta\gamma(\alpha+\epsilon)/\alpha}\Bigg\}\\
&\quad<+\infty.
\end{aligned}$$

B.1.3 Proof of L2)

By the induction hypothesis, $p_n^{(l-1)}$ converges to $p^{(l-1)}$ in distribution with respect to the weak topology. Since the limit law is degenerate (on $p^{(l-1)}$), then for every subsequence (n') there exists a subsequence (n'') such that $p_{n''}^{(l-1)}$ converges a.s. By the induction hypothesis, $p^{(l-1)}$ is absolutely continuous with respect to the Lebesgue measure. Since $|\phi|^{\alpha}$ is a.s. continuous and, by L1.1), uniformly integrable with respect to $(p_n^{(l-1)})$, then we can write the following

$$\int||\phi\circ f||^{\alpha}p_{n''}^{(l-1)}(\mathrm{d}f)\longrightarrow\int||\phi\circ f||^{\alpha}q^{(l-1)}(\mathrm{d}f)\quad a.s.$$

Thus, $n\to+\infty$

$$\int||\phi\circ f||^{\alpha}p_n^{(l-1)}(\mathrm{d}f)\xrightarrow{p}\int||\phi\circ f||^{\alpha}q^{(l-1)}(\mathrm{d}f).$$

B.1.4 Proof of L3)

Let ϵ be as in L1.1), and let $p=(\alpha+\epsilon)/\alpha$ and $q=(\alpha+\epsilon)/\epsilon$. Then $1/p+1/q=1$. Thus, by Holder inequality

$$\int||\phi\circ f||^{\alpha}\left[1-\exp\left\{-\int_{\mathbb{S}^{k-1}}|\boldsymbol{t}^T\boldsymbol{s}|^{\alpha}\left(\frac{1}{n}||\sigma_w(\phi\circ f)||^{\alpha}\delta_{\frac{\phi\circ f}{||\phi\circ f||}}\right)(\mathrm{d}\boldsymbol{s})\right\}\right]p_n^{(l-1)}(\mathrm{d}f)$$

$$\leq \left(\int ||\phi \circ f||^{\alpha p} p_n^{(l-1)}(\mathrm{d}f) \right)^{1/p}$$
$$\times \left(\int \left[1 - \exp\left\{ - \int_{\mathbb{S}^{k-1}} |\boldsymbol{t}^T \boldsymbol{s}|^\alpha \left(\frac{1}{n} ||\sigma_w(\phi \circ f)||^\alpha \delta_{\frac{\phi \circ f}{||\phi \circ f||}} \right) (\mathrm{d}\boldsymbol{s}) \right\} \right]^q p_n^{(l-1)}(\mathrm{d}f) \right)^{1/q}$$

Since we defined $p = (\alpha + \epsilon)/\alpha$ and $q = (\alpha + \epsilon)/\epsilon$, i.e. we set $q > 1$, then we can write the following

$$\left(\int ||\phi \circ f||^{\alpha p} p_n^{(l-1)}(\mathrm{d}f) \right)^{1/p}$$
$$\times \left(\int \left[1 - \exp\left\{ - \int_{\mathbb{S}^{k-1}} |\boldsymbol{t}^T \boldsymbol{s}|^\alpha \left(\frac{1}{n} ||\sigma_w(\phi \circ f)||^\alpha \delta_{\frac{\phi \circ f}{||\phi \circ f||}} \right) (\mathrm{d}\boldsymbol{s}) \right\} \right]^q p_n^{(l-1)}(\mathrm{d}f) \right)^{1/q}$$
$$\leq \sup_n \left[\left(\int ||\phi \circ f||^{\alpha + \epsilon} p_n^{(l-1)}(\mathrm{d}f) \right)^{1/p} \right]$$
$$\times \left(\int \left[1 - \exp\left\{ - \int_{\mathbb{S}^{k-1}} |\boldsymbol{t}^T \boldsymbol{s}|^\alpha \left(\frac{1}{n} ||\sigma_w(\phi \circ f)||^\alpha \delta_{\frac{\phi \circ f}{||\phi \circ f||}} \right) (\mathrm{d}\boldsymbol{s}) \right\} \right] p_n^{(l-1)}(\mathrm{d}f) \right)^{1/q}$$
$$\leq \sup_n \left[\left(\int ||\phi \circ f||^{\alpha + \epsilon} p_n^{(l-1)}(\mathrm{d}f) \right)^{1/p} \right]$$
$$\times \left(\left[\int_{\mathbb{S}^{k-1}} |\boldsymbol{t}^T \boldsymbol{s}|^\alpha \left(\frac{1}{n} ||\sigma_w(\phi \circ f)||^\alpha \delta_{\frac{\phi \circ f}{||\phi \circ f||}} \right) (\mathrm{d}\boldsymbol{s}) \right] p_n^{(l-1)}(\mathrm{d}f) \right)^{1/q} \to 0,$$

as $n \to \infty$, by L1.1).

B.1.5 Combine L1), L2) and L3)

We combine L1), L2) and L3) to prove the large n behavior of the i-th coordinate $n^{-1/\alpha} f_i(x, n)$. From (20)

$$\mathbb{E}[\mathrm{e}^{\mathrm{i}\boldsymbol{t}^T f_i^{(l)}(\boldsymbol{X}, n)}]$$
$$= \mathbb{E}\left[\exp\left\{ - \int_{\mathbb{S}^{k-1}} |\boldsymbol{t}^T \boldsymbol{s}|^\alpha \left(||\sigma_b \mathbf{1}||^\alpha \delta_{\frac{\mathbf{1}}{||\mathbf{1}||}} \right) (\mathrm{d}\boldsymbol{s}) \right\} \right]$$
$$\times \mathbb{E}\left[\left(\int \exp\left\{ - \int_{\mathbb{S}^{k-1}} |\boldsymbol{t}^T \boldsymbol{s}|^\alpha \left(\frac{1}{n} ||\sigma_w(\phi \circ f)||^\alpha \delta_{\frac{\phi \circ f}{||\phi \circ f||}} \right) (\mathrm{d}\boldsymbol{s}) \right\} p_n^{(l-1)}(\mathrm{d}f) \right)^n \right].$$
$$= \mathbb{E}\left[\exp\left\{ - \int_{\mathbb{S}^{k-1}} |\boldsymbol{t}^T \boldsymbol{s}|^\alpha \left(||\sigma_b \mathbf{1}||^\alpha \delta_{\frac{\mathbf{1}}{||\mathbf{1}||}} \right) (\mathrm{d}\boldsymbol{s}) \right\} \right]$$
$$\times \mathbb{E}\left[\mathbb{1}_{\{(p_n^{(l-1)} \in I)\}} \left(\int \exp\left\{ - \int_{\mathbb{S}^{k-1}} |\boldsymbol{t}^T \boldsymbol{s}|^\alpha \left(\frac{1}{n} ||\sigma_w(\phi \circ f)||^\alpha \delta_{\frac{\phi \circ f}{||\phi \circ f||}} \right) (\mathrm{d}\boldsymbol{s}) \right\} p_n^{(l-1)}(\mathrm{d}f) \right)^n \right].$$

Then, by Lagrange theorem, there exists a value $\theta_n \in [0, 1]$ such that the following equality holds true

$$1 - \exp\left\{ - \int_{\mathbb{S}^{k-1}} |\boldsymbol{t}^T \boldsymbol{s}|^\alpha \left(\frac{1}{n} ||\sigma_w(\phi \circ f)||^\alpha \delta_{\frac{\phi \circ f}{||\phi \circ f||}} \right) (\mathrm{d}\boldsymbol{s}) \right\}$$
$$= \left(\int_{\mathbb{S}^{k-1}} |\boldsymbol{t}^T \boldsymbol{s}|^\alpha \left(\frac{1}{n} ||\sigma_w(\phi \circ f)||^\alpha \delta_{\frac{\phi \circ f}{||\phi \circ f||}} \right) (\mathrm{d}\boldsymbol{s}) \right)$$
$$\times \exp\left\{ -\theta_n \int_{\mathbb{S}^{k-1}} |\boldsymbol{t}^T \boldsymbol{s}|^\alpha \left(\frac{1}{n} ||\sigma_w(\phi \circ f)||^\alpha \delta_{\frac{\phi \circ f}{||\phi \circ f||}} \right) (\mathrm{d}\boldsymbol{s}) \right\},$$

thus

$$\exp\left\{ - \int_{\mathbb{S}^{k-1}} |\boldsymbol{t}^T \boldsymbol{s}|^\alpha \left(\frac{1}{n} ||\sigma_w(\phi \circ f)||^\alpha \delta_{\frac{\phi \circ f}{||\phi \circ f||}} \right) (\mathrm{d}\boldsymbol{s}) \right\}$$
$$= 1 - \left(\int_{\mathbb{S}^{k-1}} |\boldsymbol{t}^T \boldsymbol{s}|^\alpha \left(\frac{1}{n} ||\sigma_w(\phi \circ f)||^\alpha \delta_{\frac{\phi \circ f}{||\phi \circ f||}} \right) (\mathrm{d}\boldsymbol{s}) \right)$$

$$\times \exp\left\{-\theta_n \int_{\mathbb{S}^{k-1}} |\boldsymbol{t}^T\boldsymbol{s}|^\alpha \left(\frac{1}{n}||\sigma_w(\phi\circ f)||^\alpha \delta_{\frac{\phi\circ f}{||\phi\circ f||}}\right)(\mathrm{d}\boldsymbol{s})\right\}$$
$$= 1 - \left(\int_{\mathbb{S}^{k-1}} |\boldsymbol{t}^T\boldsymbol{s}|^\alpha \left(\frac{1}{n}||\sigma_w(\phi\circ f)||^\alpha \delta_{\frac{\phi\circ f}{||\phi\circ f||}}\right)(\mathrm{d}\boldsymbol{s})\right)$$
$$+ \left(\int_{\mathbb{S}^{k-1}} |\boldsymbol{t}^T\boldsymbol{s}|^\alpha \left(\frac{1}{n}||\sigma_w(\phi\circ f)||^\alpha \delta_{\frac{\phi\circ f}{||\phi\circ f||}}\right)(\mathrm{d}\boldsymbol{s})\right)$$
$$\times \left(1 - \exp\left\{-\theta_n \int_{\mathbb{S}^{k-1}} |\boldsymbol{t}^T\boldsymbol{s}|^\alpha \left(\frac{1}{n}||\sigma_w(\phi\circ f)||^\alpha \delta_{\frac{\phi\circ f}{||\phi\circ f||}}\right)(\mathrm{d}\boldsymbol{s})\right\}\right).$$

Now, since

$$0 \le \int\int_{\mathbb{S}^{k-1}} |\boldsymbol{t}^T\boldsymbol{s}|^\alpha \left(\frac{1}{n}||\sigma_w(\phi\circ f)||^\alpha \delta_{\frac{\phi\circ f}{||\phi\circ f||}}\right)(\mathrm{d}\boldsymbol{s})$$
$$\times \left[1 - \exp\left\{-\theta_n \int_{\mathbb{S}^{k-1}} |\boldsymbol{t}^T\boldsymbol{s}|^\alpha \left(\frac{1}{n}||\sigma_w(\phi\circ f)||^\alpha \delta_{\frac{\phi\circ f}{||\phi\circ f||}}\right)(\mathrm{d}\boldsymbol{s})\right\}\right] p_n^{(l-1)}(\mathrm{d}f)$$
$$\le \int\int_{\mathbb{S}^{k-1}} |\boldsymbol{t}^T\boldsymbol{s}|^\alpha \left(\frac{1}{n}||\sigma_w(\phi\circ f)||^\alpha \delta_{\frac{\phi\circ f}{||\phi\circ f||}}\right)(\mathrm{d}\boldsymbol{s})$$
$$\times \left[1 - \exp\left\{- \int_{\mathbb{S}^{k-1}} |\boldsymbol{t}^T\boldsymbol{s}|^\alpha \left(\frac{1}{n}||\sigma_w(\phi\circ f)||^\alpha \delta_{\frac{\phi\circ f}{||\phi\circ f||}}\right)(\mathrm{d}\boldsymbol{s})\right\}\right] p_n^{(l-1)}(\mathrm{d}f)$$

then

$$\mathbb{E}[\mathrm{e}^{\mathrm{i}\boldsymbol{t}^T f_i^{(l)}(\boldsymbol{X},n)}]$$
$$\le \mathbb{E}\left[\exp\left\{-\int_{\mathbb{S}^{k-1}} |\boldsymbol{t}^T\boldsymbol{s}|^\alpha \left(||\sigma_b\mathbf{1}||^\alpha \delta_{\frac{\mathbf{1}}{||\mathbf{1}||}}\right)(\mathrm{d}\boldsymbol{s})\right\}\right]$$
$$\times \mathbb{E}\left[\mathbb{1}_{\{(p_n^{(l-1)}\in I)\}}\left(1 - \int\int_{\mathbb{S}^{k-1}} |\boldsymbol{t}^T\boldsymbol{s}|^\alpha \left(\frac{1}{n}||\sigma_w(\phi\circ f)||^\alpha \delta_{\frac{\phi\circ f}{||\phi\circ f||}}\right)(\mathrm{d}\boldsymbol{s}) p_n^{(l-1)}(\mathrm{d}f)\right.\right.$$
$$+ \int\int_{\mathbb{S}^{k-1}} |\boldsymbol{t}^T\boldsymbol{s}|^\alpha \left(\frac{1}{n}||\sigma_w(\phi\circ f)||^\alpha \delta_{\frac{\phi\circ f}{||\phi\circ f||}}\right)(\mathrm{d}\boldsymbol{s})$$
$$\left.\left.\times \left[1 - \exp\left\{- \int_{\mathbb{S}^{k-1}} |\boldsymbol{t}^T\boldsymbol{s}|^\alpha \left(\frac{1}{n}||\sigma_w(\phi\circ f)||^\alpha \delta_{\frac{\phi\circ f}{||\phi\circ f||}}\right)(\mathrm{d}\boldsymbol{s})\right\}\right] p_n^{(l-1)}(\mathrm{d}f)\right)^n\right].$$

Thus, by using the definition of the exponential function, i.e. $\mathrm{e}^x = \lim_{n\to+\infty}(1+x/n)^n$, and L1)-L3) we have

$$\mathbb{E}[\mathrm{e}^{\mathrm{i}\boldsymbol{t}^T f_i^{(l)}(\boldsymbol{X},n)}]$$
$$\to \exp\left\{-\int_{\mathbb{S}^{k-1}} |\boldsymbol{t}^T\boldsymbol{s}|^\alpha \left(||\sigma_b\mathbf{1}||^\alpha \delta_{\frac{\mathbf{1}}{||\mathbf{1}||}}\right)(\mathrm{d}\boldsymbol{s})\right\}$$
$$\times \exp\left\{-\int\int_{\mathbb{S}^{k-1}} |\boldsymbol{t}^T\boldsymbol{s}|^\alpha \left(||\sigma_w(\phi\circ f)||^\alpha \delta_{\frac{\phi\circ f}{||\phi\circ f||}}\right)(\mathrm{d}\boldsymbol{s}) q^{(l-1)}(\mathrm{d}f)\right\}$$

as $n\to+\infty$. That is, we proved that the large n limiting distribution of $f_i^{(l)}(x,n)$ is $\mathrm{St}_k(\alpha,\Gamma(l))$, where

$$\Gamma(l) = ||\sigma_b\mathbf{1}||^\alpha \delta_{\frac{\mathbf{1}}{||\mathbf{1}||}} + \int ||\sigma_w(\phi\circ f)||^\alpha \delta_{\frac{\phi\circ f}{||\phi\circ f||}} q^{(l-1)}(\mathrm{d}f). \tag{24}$$

C Finite-dimensional projections

We show that, as $n \to +\infty$,

$$(f_i^{(l)}(\boldsymbol{X}, n))_{i\geq 1} \xrightarrow{w} \bigotimes_{i\geq 1} \mathrm{St}_k(\alpha, \Gamma(l)), \tag{25}$$

by proving the large n asymptotic behavior of any finite linear combination of the $f_i^{(l)}(\boldsymbol{X}, n)$'s, for $i \in \mathcal{L} \subset \mathbb{N}$. See, e.g. Billingsley (1999) and reference therein. Following the notation of Matthews et al. (2018b), consider a finite linear combination of the function values without the bias. In other terms, let us consider

$$T^{(l)}(\mathcal{L}, p, \boldsymbol{X}, n) = \sum_{i\in\mathcal{L}} p_i[f_i^{(l)}(\boldsymbol{X}, n) - b_i^{(l)}\mathbf{1}].$$

Then, we write

$$\begin{aligned}
T^{(l)}(\mathcal{L}, p, \boldsymbol{X}, n) &= \sum_{i\in\mathcal{L}} p_i[f_i^{(l)}(\boldsymbol{X}, n) - b_i^{(l)}\mathbf{1}]\\
&= \sum_{i\in\mathcal{L}} p_i \left[\frac{1}{n^{1/\alpha}} \sum_{j=1}^{n} w_{i,j}^{(l)}(\phi \circ f_j^{(l-1)}(\boldsymbol{X}, n))\right]\\
&= \frac{1}{n^{1/\alpha}} \sum_{j=1}^{n} \sum_{i\in\mathcal{L}} p_i w_{i,j}^{(l)}(\phi \circ f_j^{(l-1)}(\boldsymbol{X}, n))\\
&= \frac{1}{n^{1/\alpha}} \sum_{j=1}^{n} \gamma_j^{(l)}(\mathcal{L}, p, \boldsymbol{X}, n),
\end{aligned}$$

where

$$\gamma_j^{(l)}(\mathcal{L}, p, \boldsymbol{X}, n) = \sum_{i\in\mathcal{L}} p_i w_{i,j}^{(l)}(\phi \circ f_j^{(l-1)}(\boldsymbol{X}, n)).$$

Then,

$$\begin{aligned}
&\varphi_{T^{(l)}(\mathcal{L},p,\mathbf{X},n)\,|\,\{f_j^{(l-1)}(\boldsymbol{X},n)\}_{j=1,\dots,n}}(\boldsymbol{t})\\
&= \mathbb{E}[e^{\mathrm{i}\boldsymbol{t}^T T^{(l)}(\mathcal{L},p,\boldsymbol{X},n)} \mid \{f_j^{(l-1)}(\boldsymbol{X}, n)\}_{j=1,\dots,n}]\\
&= \mathbb{E}\left[\exp\left\{\mathrm{i}\boldsymbol{t}^T \left[\frac{1}{n^{1/\alpha}} \sum_{j=1}^{n}\sum_{i\in\mathcal{L}} \alpha_i w_{i,j}^{(l)}(\phi \circ f_j^{(l-1)}(\boldsymbol{X}, n))\right]\right\} \mid \{f_j^{(l-1)}(\boldsymbol{X}, n)\}_{j=1,\dots,n}\right]\\
&= \prod_{j=1}^{n}\prod_{i\in\mathcal{L}} \mathbb{E}\left[\exp\left\{\mathrm{i}\boldsymbol{t}^T \frac{1}{n^{1/\alpha}} p_i w_{i,j}^{(l)}(\phi \circ f_j^{(l-1)}(\boldsymbol{X}, n)) \mid \{f_j^{(l-1)}(\boldsymbol{X}, n)\}_{j=1,\dots,n}\right\}\right]\\
&= \prod_{j=1}^{n}\prod_{i\in\mathcal{L}} \mathrm{e}^{-\frac{p_i^\alpha \sigma_w^\alpha}{n}|\boldsymbol{t}^T(\phi\circ f_j^{(l-1)}(\boldsymbol{X},n))|^\alpha}\\
&= \exp\left\{-\int_{\mathbb{S}^{k-1}} |\boldsymbol{t}^T\boldsymbol{s}|^\alpha \left(\frac{1}{n}\sum_{j=1}^{n}\sum_{i\in\mathcal{L}} ||p_i\sigma_w(\phi\circ f_j^{(l-1)}(\boldsymbol{X}, n))||^\alpha \delta_{\frac{\phi\circ f_j^{(l-1)}(\boldsymbol{X},n)}{||\phi\circ f_j^{(l-1)}(\boldsymbol{X},n)||}}\right)(\mathrm{d}\boldsymbol{s})\right\}
\end{aligned}$$

That is,

$$T^{(l)}(\mathcal{L}, p, \boldsymbol{X}, n) \mid \{f_j^{(l-1)}(\boldsymbol{X}, n)\}_{j=1,\dots,n} \overset{\mathrm{d}}{=} \boldsymbol{S}_{\alpha,\Gamma^{(l)}}.$$

where

$$\Gamma^{(l)} = \frac{1}{n}\sum_{j=1}^{n}\sum_{i\in\mathcal{L}} ||p_i\sigma_w(\phi\circ f_j^{(l-1)}(\boldsymbol{X}, n))||^\alpha \delta_{\frac{\phi\circ f_j^{(l-1)}(\boldsymbol{X},n)}{||\phi\circ f_j^{(l-1)}(\boldsymbol{X},n)||}}.$$

Then, along lines similar to the proof of the large n asymptotics for the i-th coordinate, we have

$$\mathbb{E}[\mathrm{e}^{\mathrm{i}\boldsymbol{t}^{T} T^{(l)}(\mathcal{L},p,\boldsymbol{X},n)}] \\ \to \exp\left\{-\int\int_{\mathbb{S}^{k-1}} |\boldsymbol{t}^{T}\boldsymbol{s}|^{\alpha}\left(\sum_{i\in\mathcal{L}} ||p_i\sigma_w(\phi\circ f)||^{\alpha}\delta_{\frac{\phi\circ f}{||\phi\circ f||}}\right)(\mathrm{d}\boldsymbol{s})q^{(l-1)}(\mathrm{d}f)\right\}$$

as $n\to+\infty$. This complete the proof.

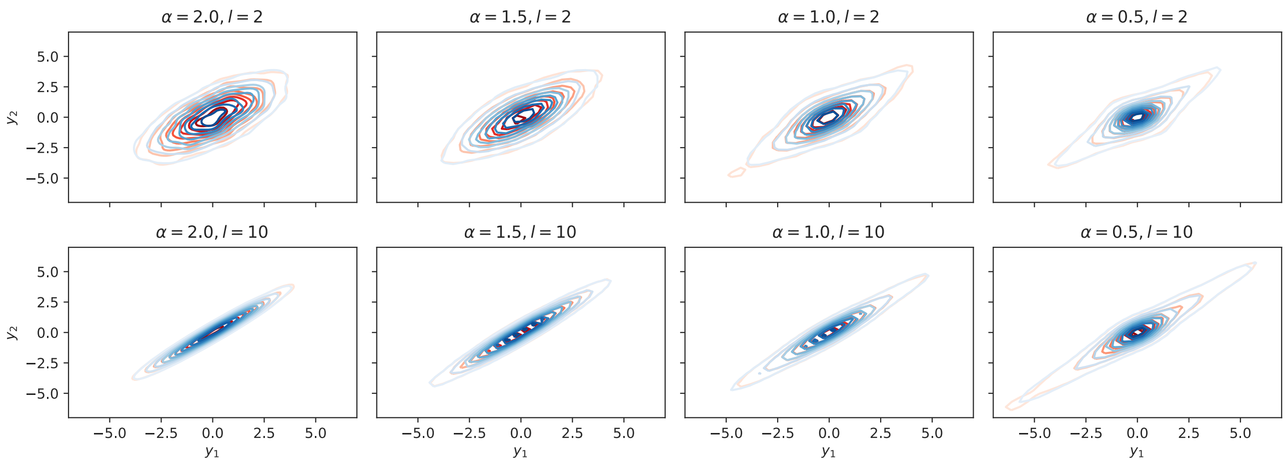

Figure 1: 2D-KDE estimates for $y \sim f_1^{(l)}(x, x', n)$ (red) and $y \sim \mathrm{St}_k(\alpha, \widetilde{\Gamma}(x, x', l, M))$ (blue).

D Numerical evaluation of the recursion

In this section we perform a preliminary numerical investigation of the approach proposed in Section 6.1 for the evaluation of recursion (13)-(14). We consider only the case of two inputs $x = -0.5, x' = 1.0$ (i.e. a bivariate stable distribution) and we use pseudo random numbers, i.e. standard Monte Carlo (MC), instead of quasi random numbers as suggested in the main text. We consider $\sigma_b = \sigma_w = 1$, the tanh activation function, different values of the stability index α, and both shallow ($l = 2$, i.e. 1 hidden layer) and deep ($l = 10$) NNs. In all cases the networks are wide: $n = 300$. In Figure 1 we compare the bivariate distributions of: i) the first dimension ($i = 1$) of the NN distribution $y \sim f_1^{(l)}(x, x', n)$ ii) its asymptotic distribution $y \sim \mathrm{St}_k(\alpha, \widetilde{\Gamma}(x, x', l, M))$ as $n \to +\infty$. In ii) we use $M = 1000$ MC samples to evaluate the discrete spectral measure $\widetilde{\Gamma}$ at each layer. In both i) and ii) we generate 100.000 samples for $y \in \mathbb{R}^2$ that are used to obtain the 2D-KDE plots of Figure 1. We can observe close agreement in all cases considered (the "squarish" level curves near the central regions for small α are an artifact due to the specific KDE estimation algorithm employed and its non-robustness to "outliers").

The code at `https://github.com/stepelu/deep-stable` contains a `numpy`-based Python implementation of the algorithms used for the simulation of scalar and multivariate stable distributions. Scalar stable variables are generated according to Weron (1996); Weron et al. (2010). In the case of multivariate stable variables the algorithm implemented is the one reported in Nolan (2008), note that the discrete spectral measure needs to be symmetrized. The code also contains the routines used to sample from $f^{(l)}(x, n)$ and from $\mathrm{St}_k(\alpha, \widetilde{\Gamma}(x, l, M))$. The implementation does not rely on advanced features so it is easily portable to deep learning frameworks such as `tensorflow` or `pytorch`. By modifying the calls to uniform random generators, it is also possible to use quasi random number generators. In all cases, the usual precaution to exclude the extremes of the supports of the uniform distributions involved (i.e. to sample from $\mathcal{U}(0, 1)$, not from $\mathcal{U}[0, 1]$) applies.

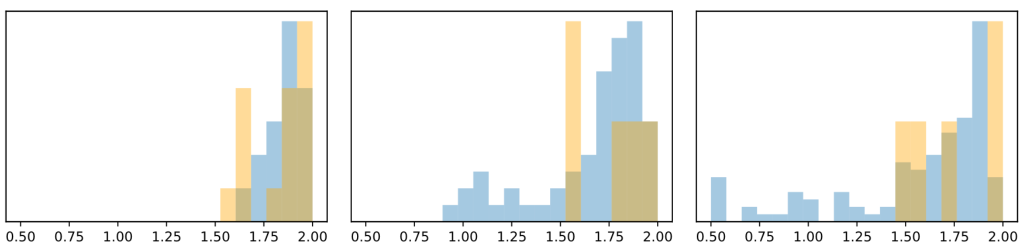

Figure 2: Histograms for the stability indexes α of the fitted Stable distributions for all layers, in blue for the weights and in yellow for the biases. The models, from left-to-right, are: VGG-16, ResNet-50 and ResNet-101.

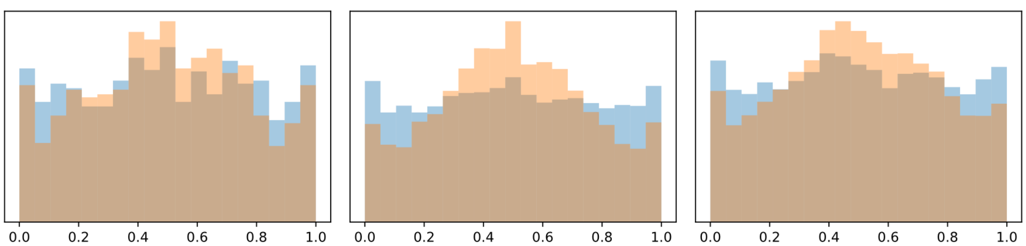

Figure 3: Histograms of the cdf evaluations for the Stable distribution (blue) and Gaussian distribution (orange) fitted to the weights of the first three layers (left-to-right) of the VGG16 model.

E Empirical analysis of trained CNNs

In this section we investigate whether trained models exhibit parameters distributions close to that of Stable distributions with stability index $0 < \alpha < 2$, i.e. non-Gaussian. We consider 3 models from the PyTorch's TorchVision repository, i.e. CNNs trained on ImageNet. While fully connected networks are ideal starting points for a theoretical analysis, it seems possible to expand our results to CNNs as done in Garriga-Alonso et al. (2019) for Gaussian Processes (GP). This allows us to investigate the parameter distributions of trained model in the "realistic" setting of overparametrized models applied to big datasets with the use of batch normalization and adaptive optimizers.

We restrict our analysis to marginal distributions and for each layer we collect all weights (CNN filters) and biases and fit a Stable distribution via maximum likelihood estimation (MLE). In Figure 2 we plot histograms for the stability indexes α of the fitted Stable distributions for all layers. We see that distributions are often non-Gaussian, and α seems to be decreasing with the depth of the model. However, it is not possible to draw definitive conclusions from this short experiment.

To obtain an indication of the goodness of fit of Stable distributions to the parameters, for the first three layers of VGG-16 ($\alpha \sim 1.7$) we: fit a Stable distribution to the weights; compute the cumulative distribution function (cdf) of this Stable distribution for each weight; fit a Gaussian distribution to the weights; compute the cdf of this Gaussian distribution for each weight; plot in Figure 3 a histogram of the cdf evaluations for the Stable and Gaussian distribution. In case of perfect fit the histogram should be flat, as the cdf evaluations should be iid uniformly distributed. We see that the fit of the Stable distributions is as expected better than the fit of Gaussian distributions, especially in the tails. The peculiar behavior at extremes of the histograms (tails) could be due to the use of truncated initializations in PyTorch. We validated MLE (limited here to $\alpha > 0.5$) and cdf computation on synthetic data generated via `sample_stable()` from the code accompanying this paper.