Infinite-channel deep convolutional Stable neural networks

Daniele Bracale
Department of Statistics
University of Michigan
dbracale@umich.edu

Stefano Favaro
Department ESOMAS
University of Torino and
Collegio Carlo Alberto
stefano.favaro@unito.it

Sandra Fortini
Department of Decision Sciences
Bocconi University
sandra.fortini@unibocconi.it

Stefano Peluchetti
Cogent Labs
speluchetti@cogent.co.jp

Abstract

The connection between infinite-width neural networks (NNs) and Gaussian processes (GPs) is well known since the seminal work of Neal [1996]. While numerous theoretical refinements have been proposed in recent years, the connection between NNs and GPs relies on two critical distributional assumptions on the NN's parameters: i) finite variance ii) independent and identical distribution (iid). In this paper, we consider the problem of removing assumption i) in the context of deep feed-forward convolutional NNs. We show that the infinite-channel limit of a deep feed-forward convolutional NNs, under suitable scaling, is a stochastic process with multivariate stable finite-dimensional distributions, and we give an explicit recursion over the layers for their parameters. Our contribution extends recent results of Favaro et al. [2020] to convolutional architectures, and it paves the way to exciting lines of research that rely on GP limits.

1   Introduction

Fully-connected NNs are defined by an interleaved application of affine transforms and non-linear functions evaluated element-wise. Associating a distribution to the parameters of a NN allows us to consider the NN as a probabilistic model. Modern NNs typically operate in the over-parametrized regime, with millions of parameters representing a standard setting. As a way forward, Neal [1996] established the equivalence between a certain class of shallow probabilistic NNs and corresponding limiting GPs when the NN's width, hence the dimensionality of its parameters, becomes infinite. Recently, the NN-GP correspondences has been extended to deep fully-connected Lee et al. [2018], Matthews et al. [2018] and convolutional Novak et al. [2018], Garriga-Alonso et al. [2018] NNs. Two common properties of the NN's parameters' distribution underlay the NN-GP correspondence: i) finite variance ii) iid distribution. In this paper we extend the results of Novak et al. [2018], Garriga-Alonso et al. [2018] to iid parameters distributed according to a Stable distribution (SD), effectively removing assumption i). More precisely, we study the infinite-channel limit of CNNs in the following general setting: i) the CNN is deep, namely is composed of multiple layers; ii) biases and scaled weights are iid according to a centered symmetric SD; iii) number of convolutional channels in each network's layers goes to infinity jointly on the layers; iv) the convergence in distribution is established jointly for multiple inputs, namely the convergence concerns the class of finite dimensional distributions of the CNN viewed as a stochastic process. The use of SDs, which includes the Gaussian distribution as special case, is natural in this setting. Indeed, SDs are the most

general distribution class toward which infinite sums of iid random variables can converge in law Samoradnitsky and Taqqu [1994]. Through this paper we define a deep CNN (DCNN) of the form $f^{(1)} = W^{(1)} * x + b^{(1)}$ and $f^{(l)} = W^{(L)} * \phi(f^{(l-1)}) + b^{(l)}$ for $l = 2, \ldots, L$, where l indexes the L layers, $\{W^{(l)}\}_{l=1}^{L}$ are the weights, $\{b^{(l)}\}_{l=1}^{L}$ are the biases and ϕ is an activation function. We show that the infinite-channel limit of the DCNN, under suitable scaling on the weights, is a stochastic process whose finite-dimensional distributions are multivariate SDs Samoradnitsky and Taqqu [1994]. This process is referred to as the convolutional stable process (CSP). Our result contributes to the theory of DCNNs, and it paves the way to extend the research directions that rely on Gaussian infinite wide limits.

The paper is structured as follows. Section 2 introduces the notation and definitions. In Section 3 we define a DCNN jointly on K distinct inputs and we specify the distribution assumptions on the model parameters. In Section 4 we compute the limiting distributions jointly over K inputs and we establish the distribution, again over K inputs, arising from the projection of all spatial features to a single output vector (i.e. a readout layer). In Section 5 we conclude.

2   Notation and definitions

Denote: $[n]$ the set $\{1, \ldots, n\}$ $\forall n \in \mathbb{N}$; fixed a size $S \in \mathbb{N}$, an S-tensor of dimension $\mathbf{D}_S :=$ $D_1 \times \cdots \times D_S$ (where $D_j \in \mathbb{N}$ $\forall j \in [S]$), is an element $A \in \mathbb{R}^{\mathbf{D}_S}$; $[\mathbf{D}_S] := [D_1] \times \cdots \times [D_S]$; $|\mathbf{D}_S| := \prod_{s=1}^{S} D_s$; $A_d \in \mathbb{R}$ (where $d = (d_1, \ldots, d_S) \in [\mathbf{D}_S]$) is the component of A at position d; the norm $\|A\|^\alpha = \sum_{d \in [\mathbf{D}_S]} A_d^\alpha$; $A_{(d_1,;)}$ is the $(S-1)$-tensor of dimension $1 \times D_2 \times \cdots \times D_S$ consisting on the d_1-th position of the first dimension of A and all the other positions of A; an integral in $\mathrm{d}(A_{\{d \in [\mathbf{D}_S]\}})$ is an integral in the flattened tensor A, i.e. with $|\mathbf{D}_S|$ variables of integration; $\mathbb{1}_{(\mathbf{D}_S)}$ is the tensor of all 1s in $\mathbb{R}^{\mathbf{D}_S}$ and $\mathbb{1}_{(\mathbf{D}_S)}[d]$ the one with 1 in the d-th entry and zero otherwise; let $A, B \in \mathbb{R}^{\mathbf{E}_S}$, $A \otimes B = \sum_{e \in [\mathbf{E}]} A_e B_e \in \mathbb{R}$ is the *Frobenius product*; let $(A, B) \in \mathbb{R}^{\mathbf{D}_S \times \mathbf{E}_{S'}} \times \mathbb{R}^{\mathbf{E}_{S'}}$, $A \square_\diamondsuit B \in \mathbb{R}^{\mathbf{D}_S}$ is called *square product under* $\diamondsuit$(where $\diamondsuit$ operates within tensors of the same size into $\mathbb{R}$, e.g. the Frobenius product), where each position $d \in [\mathbf{D}_S]$ of $A \square_\diamondsuit B$ is $A_{(d,;)} \diamondsuit B \in \mathbb{R}$; $\square$ is simply $\square_\otimes$ and called *square product*; let $(A, B) \in \mathbb{R}^{\mathbf{D}_S} \times \mathbb{R}^{\mathbf{E}_{S'}}$, $A \triangle B = (AB_b)_{\{b \in [\mathbf{E}_{S'}]\}} \in \mathbb{R}^{\mathbf{D}_S \times \mathbf{E}_{S'}}$ is called *bias product*. When we specify some dimensions over an operation, it means that the operation is applied through *all dimensions except for* the specified ones, e.g. in $\overset{(P,K)}{\square}$, $\square$ is applied through all dimensions except for dimensions P and K. When dimensions are under the operation, the operation is applied *only to the specified dimensions*; $A \sim \mathrm{St}_{\mathbf{D}_S}(\alpha, \Gamma)$ indicates that the corresponding flattened vector has characteristic function $\varphi_A(t) := \mathbb{E}[e^{\mathrm{i}\, \mathbf{t} \otimes A}] = \exp\{-\int_{\mathbb{S}^{|\mathbf{D}_S|-1}} |\mathbf{t} \otimes \mathbf{s}|^\alpha \Gamma(\mathrm{d}\, \mathbf{s})\}$ for all $t \in \mathbb{R}^{\mathbf{D}_S}$, where Γ is a spectral measure on $\mathbb{S}^{|\mathbf{D}_S|-1} = \{z \in \mathbb{R}^{|\mathbf{D}_S|} : \|z\| = 1\}$ (more details on "SM: Stable random variables").

3   Stable convolutional networks

Shallow CNN. The input to a convolution is a tensor $x \in \mathbb{R}^{C \times \mathbf{P}_S}$ where S is the number of spacial dimensions, P_s is the size of the s-th spacial dimension $\forall s \in [S]$ and C is the number of channels. The defining property of a convolution is that the same collection of filters (weights) is applied to multiple patches extracted from the input tensor x. The filter size must thus agree with the extracted patches sizes. There is great flexibility in defining the specific details of a given convolutional transform, including its striding, padding, and dilation characteristics. See Dumoulin and Visin [2016] for a comprehensive account. In this paper we consider the following general setting. We define the filters $W \in \mathbb{R}^{C' \times C \times \mathbf{G}_S}$ where G_s ($\leq P_s$ and the equality makes the network a S-dimensional fully connected NN) is the filter size across the s-th space dimension $\forall s \in [S]$ and C' is the number of output channels. We also define a bias term $b \in \mathbb{R}^{C'}$. A convolution transform over x results in an output tensor $y \in \mathbb{R}^{C' \times \mathbf{P}'_S}$ where P'_s is the size of the s-th spacial output dimension. P'_s depends on both P_s and on the convolution characteristics. We write $y = W * x + b$, where the product $*$ is defined as follows. A patch extracted from x at output position p is $x_{\star p} \in \mathbb{R}^{C \times \mathbf{G}_S}$, $x_{\star p} = x_{1:C, \star p}$ where $\star : p \mapsto \star p$ is a function that depends on the moving window chosen in the structure and returns the positions of x associated with the corresponding output position p. When extraction happens outside of x, i.e. when an input position i is such that $i < 1$ or $i > P$, the padded values

(often a constant) are taken as input. The convolution transform at output position p is thus given by $y_p = W x_{\star p} + b \in \mathbb{R}^{C'}$. Finally $y \in \mathbb{R}^{C'}$ is obtained by stacking y_p over the $\mathbf{P}'_S$ output positions, i.e. the NN can be rewritten as

$$y = [W x_{\star p} + b]_{\{p\}} \tag{1}$$

DCNN with K inputs. We extend the definition (1). A DCNN is defined by multiple layers of convolutional transforms followed by the application of an element-wise activation function ϕ. We consider the case where all layers have the same number of channels $C^{(1)} = \cdots = C^{(L)} = C$. Define $\mathbf{P}^{(l)} = \mathbf{P}^{(l)}_{S^{(l)}}$ for $l \in [L] \cup \{0\}$ and $\mathbf{G}^{(l)} = \mathbf{G}^{(l)}_{S^{(l-1)}}$ for $l \in [L]$. Consider K inputs $x^{(k)} \in \mathbb{R}^{C^{(0)} \times \mathbf{P}^{(0)}}$, $k \in [K]$ and define $x^{(1:K)} = (x^{(1)}, \ldots, x^{(K)})^T \in \mathbb{R}^{C^{(0)} \times \mathbf{P}^{(0)} \times K}$. For each $x^{(k)}$ the convolutional structure remains constant, that is the filters and the bias terms do not depend on k. Since they do not even depend on the positions, a DCNN of L layers with with K inputs can be defined as follows

$$\begin{cases} f^{(0)(1:K)} = x^{(1:K)} & \in \mathbb{R}^{C^{(0)} \times \mathbf{P}^{(0)} \times K} \\ f^{(1)(1:K)} = f^{(l)}(x^{(1:K)}) = W^{(1)} \overset{(\mathbf{P}^{(1)}, K)}{\boxdot} x^{(1:K)}_\star + b^{(1)} \triangle \mathbb{1}_{(\mathbf{P}^{(1)} \times K)} & \in \mathbb{R}^{\infty \times \mathbf{P}^{(1)} \times K} \\ f^{(l)(1:K)} = f^{(l)}(x^{(1:K)}, C) = \frac{1}{C^{1/\alpha}} W^{(l)} \overset{(\mathbf{P}^{(l)}, K)}{\boxdot} \phi(f^{(l-1)(1:K)}_\star) \\ \quad + b^{(l)} \triangle \mathbb{1}_{(\mathbf{P}^{(l)} \times K)} & \in \mathbb{R}^{\infty \times \mathbf{P}^{(l)} \times K} \end{cases} \tag{2}$$

where the last holds for $l = 2, \ldots, L$ and where $\alpha \in (0,2]$, $W^{(l)} \in \mathbb{R}^{\infty \times C^{(l-1)} \times \mathbf{G}^{(l)}}$, $b^{(l)} \in \mathbb{R}^\infty$ for each $l \in [L]$, and $x^{(1:K)}_\star = [x^{(k)}_{\star p^{(1)}}]_{\{(p^{(1)}, k) \in [\mathbf{P}^{(1)} \times K]\}} \in \mathbb{R}^{C^{(0)} \times \mathbf{G}^{(1)} \times \mathbf{P}^{(1)} \times K}$ and $f^{(l-1)(1:K)}_\star = [f^{(l-1)(k)}_{\star p^{(l)}}]_{\{(p^{(l)}, k) \in [\mathbf{P}^{(l)} \times K]\}} \in \mathbb{R}^{C^{(l-1)} \times \mathbf{G}^{(l)} \times \mathbf{P}^{(l)} \times K}$ for $l = 2, \ldots L$, where $f^{(l-1)}_{\star p^{(l)}} := f^{(l-1)}_{(1:C^{(l-1)}, \star p^{(l)})} \in \mathbb{R}^{C^{(l-1)} \times \mathbf{G}^{(l)}}$ is a patch extracted from $f^{(l-1)}$ at output position $p^{(l)} \in [\mathbf{P}^{(l)}]$. We defer to "SM: the patch operator" for a more detailed explanation.

Assumptions on the parameters and on the activation function.

H1) $\forall l \in [L], c^{(l)} \geq 1, g^{(l)} \in \mathbf{G}^{(l)}: \quad W^{(l)}_{(c^{(l)}, c^{(l-1)}, g^{(l)})} \overset{iid}{\sim} \text{St}(\alpha, \sigma_w) \quad iid \quad b^{(l)}_{c^{(l)}} \overset{iid}{\sim} \text{St}(\alpha, \sigma_b) \tag{3}$

H2) $\phi : \mathbb{R} \to \mathbb{R}$ with finite discontinuities: $\forall s \in \mathbb{R} \ |\phi(s)| \leq a + b|s|^\beta$ some $a, b > 0$ and $\beta < 1$ (4)

4 Main theorems: infinitely wide limits

We study the limiting distribution of $f^{(l)(1:K)} = f^{(l)}(x^{(1:K)}, C)$ as $C \to \infty$, $\forall l \in [L]$. Let $f^{(l)(1:K)}_\infty \in \mathbb{R}^{\infty \times \mathbf{P}^{(l)} \times K}$ be this limit (i.e. the joint limit random variable over all (infinite) channels, positions and inputs). By the Cramér-Wold theorem it is sufficient to prove the large C asymptotic behavior of any linear combination of $f^{(l)(1:K)}_{(c^{(l)}, :)}$'s (see, e.g. Billingsley [1999] for details), where $f^{(l)(1:K)}_{(c^{(l)}, :)}$ is the $c^{(l)}$-th channel of $f^{(l)(1:K)}$, and, from (2), it can be rewritten as

$$\begin{cases} f^{(1)(1:K)}_{(c^{(1)}, :)} = W^{(1)}_{(c^{(1)}, :, :)} \overset{(\mathbf{P}^{(1)}, K)}{\otimes} x^{(1:K)}_\star + b^{(1)}_{c^{(1)}} \mathbb{1}_{(\mathbf{P}^{(1)} \times K)} & \in \mathbb{R}^{\mathbf{P}^{(1)} \times K} \\ f^{(l)(1:K)}_{(c^{(l)}, :)} = \frac{1}{C^{1/\alpha}} W^{(l)}_{(c^{(l)}, :, :)} \overset{(\mathbf{P}^{(l)}, K)}{\otimes} \phi(f^{(l-1)(1:K)}_\star) + b^{(l)}_{c^{(l)}} \mathbb{1}_{(\mathbf{P}^{(l)} \times K)} & \in \mathbb{R}^{\mathbf{P}^{(l)} \times K}, l = 2, \ldots, L \end{cases} \tag{5}$$

Note: fully connected NNs are special cases when $P^{(l)} = 1 \ \forall l \in [L] \cup \{0\}$ and the patch extraction corresponds to the whole input for each convolutional transform. Thus we prove: **Theorem 1** by fixing $c^{(L)} \geq 1, l \in [L]$ and computing the limit distribution of $f^{(l)(1:K)}_{(c^{(l)}, :)}$ as $C \to \infty$ and **Theorem 2** by applying the Cramér-Wold theorem. We denoted with $\overset{d}{\to}$, $\overset{p}{\to}$ and $\overset{a.s.}{\to}$ respectively the convergence in distribution, in probability and almost surely. Our proof is an alternative to the Strong Law of Large Numbers for Stable random variables. The key point of the proof lies in recognizing the exchangeability of the sequence $(f^{(l)(1:K)}_{(c^{(l)}, :)})_{c^{(l)} \geq 1}$ which allows us to apply the de Finetti theorem.

First, define for each $l \in [L]$ the function $\Psi^{(l)} : \mathbb{R}^{\mathbf{P}^{(l)} \times K} \to \mathbb{R}$, $\Psi^{(l)}(z) := \frac{1}{2}\delta(z/\|z\|) + \frac{1}{2}\delta(-z/\|z\|)$ if $z \neq 0$ and 0 otherwise, where δ denotes the Dirac delta function. Recall that we don't compute the limit for the layer $l = 1$, because $f^{(1)(1:K)}_{(c^{(1)}, :)}$ is referred only to the $C^{(0)}$ channels of the input layer.

Theorem 1. *[SM A] For each* $l = 2, \ldots, L$, $f_{(c^{(l)},:)}^{(l)(1:K)} \xrightarrow{d} f_{\infty(c^{(l)},:)}^{(l)(1:K)} \sim St_{\boldsymbol{P}^{(l)} \times K}(\alpha, \Gamma_\infty^{(l)})$ *as* $C \to \infty$, *being* $\Gamma_\infty^{(l)}$ *equal to*

$$\|\sigma_b \mathbb{1}_{(\boldsymbol{P}^{(1)} \times K)}\|^\alpha \Psi^{(l)}\Big(\mathbb{1}_{(\boldsymbol{P}^{(l)} \times K)}\Big) + \int \sum_{g^{(l)} \in [\boldsymbol{G}^{(l)}]} \|\sigma_\omega \phi(f_{g^{(l)}})\|^\alpha \Psi^{(l)}\Big(\phi(f_{g^{(l)}})\Big) q^{(l-1)}(\mathrm{d} f_{\{g^{(l)} \in [\boldsymbol{G}^{(l)}]\}})$$

where $f_{g^{(l)}} \in \mathbb{R}^{\boldsymbol{P}^{(l)} \times K}$ *for each* $g^{(l)} \in \boldsymbol{G}^{(l)}$, $q^{(l-1)} = St_{\boldsymbol{P}^{(l-1)} \times K}(\alpha, \Gamma_\infty^{(l-1)})$, *and*

$$\Gamma_\infty^{(1)} = \Gamma^{(1)} = \|\sigma_b \mathbb{1}_{(\boldsymbol{P}^{(1)} \times K)}\|^\alpha \Psi^{(1)}(\mathbb{1}_{(\boldsymbol{P}^{(1)} \times K)})$$
$$+ \sum_{(c^{(0)},g^{(1)}) \in [C^{(0)} \times \boldsymbol{G}^{(1)}]} \|\sigma_\omega (x_\star^{(1:K)})_{(c^{(0)},g^{(1)})}\|^\alpha \Psi^{(1)}\Big((x_\star^{(1:K)})_{(c^{(0)},g^{(1)})}\Big)$$

Theorem 2. *[SM B] For each* $l \in [L]$, $f^{(l)(1:K)} = f^{(l)}(x^{(1:K)}, C) \xrightarrow{d} f_\infty^{(l)(1:K)} \sim \bigotimes_{c^{(l)}=1}^\infty St_{\boldsymbol{P}^{(l)} \times K}(\alpha, \Gamma_\infty^{(l)})$ *as* $C \to \infty$, *where the symbol* $\bigotimes$ *denotes the product measure.*

Readout layer on positions. Fix $l = L$. We found that $f^{(L)(1:K)} \xrightarrow{d} f_\infty^{(L)(1:K)}$, i.e. a convergence of a sequence of $\mathbb{R}^{\infty \times \boldsymbol{P}^{(L)} \times K}$-valued random variables. To gather information on the positions $\mathbf{P}^{(L)}$ we consider a linear combination with respect to $\mathbf{P}^{(L)}$, i.e. we project the $\infty \times \mathbf{P}^{(L)} \times K$ dimensional vector $f^{(L)(1:K)} = f^{(L)}(x^{(1:K)}, C)$ into one $\infty \times K$ dimensional, and we take the limit as $C \to \infty$. For a detailed explanation we defer to the "SM C".

5 Conclusions

We showed that an infinite-channel DCNN with scaled Stable parameters defines a stochastic process whose finite-dimensional distributions are multivariate SDs. The finite-dimensional distributions can be evaluated via an explicit recursion over the layers of the NN. We also established the finite-dimensional distributions arising from the NN readout layer. Several interesting theoretical developments are possible. Firstly, the results provided so far constitute the first step in establishing an NTK limit Arora et al. [2019] arising from Stable distributed parameters in convolutional architectures. Secondly, all the established convergence results are limited to the finite dimensional distributions of the NN layers. This is not enough to guarantee the convergence of the NN seen as a random function of the input space, i.e. to establish a functional limit. Further effort is thus needed to extend the results of Bracale et al. [2021] to the Stable and convolutional setting for $0 < \alpha < 2$. Doing so will also provide estimates on the (reduced, compared to the Gaussian case) smoothness proprieties of the limiting stochastic processes. Finally, it is necessary to devise efficient inference algorithms that allows to apply the stochastic processes introduced in this paper to current computer vision problems.

References

S. Arora, S. S. Du, W. Hu, Z. Li, R. Salakhutdinov, and R. Wang. On exact computation with an infinitely wide neural net. *Advances in Neural Information Processing Systems*, 32, 2019.

P. Billingsley. *Convergence of Probability Measures*. Wiley-Interscience, 2nd edition, 1999.

D. Bracale, S. Favaro, S. Fortini, and S. Peluchetti. Large-width functional asymptotics for deep gaussian neural networks. In *International Conference on Learning Representations*, 2021.

V. Dumoulin and F. Visin. A guide to convolution arithmetic for deep learning. *arXiv preprint arXiv:1603.07285*, 2016.

S. Favaro, S. Fortini, and S. Peluchetti. Stable behaviour of infinitely wide deep neural networks. In *International Conference on Artificial Intelligence and Statistics*, pages 1137–1146. PMLR, 2020.

A. Garriga-Alonso, C. E. Rasmussen, and L. Aitchison. Deep convolutional networks as shallow gaussian processes. In *International Conference on Learning Representations*, 2018.

J. Lee, Y. Bahri, R. Novak, S. S. Schoenholz, J. Pennington, and J. Sohl-Dickstein. Deep neural networks as gaussian processes. In *International Conference on Learning Representations*, 2018.

A. G. d. G. Matthews, M. Rowland, J. Hron, R. E. Turner, and Z. Ghahramani. Gaussian process behaviour in wide deep neural networks. *arXiv preprint arXiv:1804.11271*, 2018.

R. M. Neal. Bayesian learning for neural networks. *Lecture Notes in Statistics*, 1996.

R. Novak, L. Xiao, Y. Bahri, J. Lee, G. Yang, J. Hron, D. A. Abolafia, J. Pennington, and J. Sohl-dickstein. Bayesian deep convolutional networks with many channels are gaussian processes. In *International Conference on Learning Representations*, 2018.

G. Samoradnitsky and M. S. Taqqu. *Stable non-Gaussian random processes: stochastic models with infinite variance*. Routledge, 1994.

SUPPLEMENTARY MATERIALS

SM: Stable random variables

Fix $0 < \alpha \leq 2$ and $\sigma > 0$ and define the following distributions.

Definition 1. *A $\mathbb{R}$-valued random variable A is distributed as a SD with index α, skewness $\tau \in [-1, 1]$, scale σ and shift $\mu \in \mathbb{R}$, and we write $A \sim St(\alpha, \tau, \sigma, \mu)$, if the characteristic function is*

$$\varphi_A(t) = \mathbb{E}[e^{\mathrm{i}\,tA}] = e^{\psi(t)}, \quad t \in \mathbb{R} \quad \text{where}$$

$$\psi(t) = \begin{cases} -\sigma^\alpha |t|^\alpha [1 + \mathrm{i}\,\tau \tan(\frac{\alpha\pi}{2}) sign(t)] + \mathrm{i}\,\mu t & \alpha \neq 1 \\ -\sigma |t|[1 + \mathrm{i}\,\tau \frac{2}{\pi} sign(t) \log(|t|)] + \mathrm{i}\,\mu t & \alpha = 1 \end{cases}$$

As the property 1.2.16 of Samoradnitsky and Taqqu [1994] shows if $A \sim St(\alpha, \tau, \sigma, \mu)$ with $0 < \alpha < 2$ then $\mathbb{E}[|A|^r] < \infty$ for $0 < r < \alpha$, and $\mathbb{E}[|A|^r] = \infty$ for $r \geq \alpha$.

Definition 2. *A $\mathbb{R}$-valued random variable A is distributed as the symmetric α-stable distribution with scale parameter σ, and we write $A \sim St(\alpha, \sigma)$, if it is stable with $\tau = \mu = 0$, i.e. the characteristic function is*

$$\varphi_A(t) = \mathbb{E}[e^{\mathrm{i}\,tA}] = e^{-\sigma^\alpha |t|^\alpha}, \quad t \in \mathbb{R}$$

By property 1.2.3 of Samoradnitsky and Taqqu [1994], $a St(\alpha, \sigma) = \sigma St(\alpha, |a|\sigma)$ for every $a \in \mathbb{R}$.

Definition 3. *A $\mathbb{R}^D$-valued random vector A is distributed as the symmetric D-dimensional α-stable distribution with scale (finite) spectral measure Γ on $\mathbb{S}^{D-1} = \{z \in \mathbb{R}^D : \|z\| = 1\}$, and we write $A \sim St_D(\alpha, \Gamma)$, if the characteristic function is*

$$\varphi_A(\boldsymbol{t}) = \mathbb{E}[e^{\mathrm{i}\langle \boldsymbol{t}, A\rangle}] = e^{-\int_{\mathbb{S}^{D-1}} |\langle \boldsymbol{t}, \boldsymbol{s}\rangle|^\alpha \Gamma(\mathrm{d}\boldsymbol{s})}, \quad t \in \mathbb{R}^D$$

Definition 4. *A $\mathbb{R}^{\boldsymbol{D}_S}$-valued random S-tensor A is distributed as the symmetric $\boldsymbol{D}_S$-dimensional α-stable distribution with scale (finite) spectral measure Γ on $\mathbb{S}^{|\boldsymbol{D}_S|-1} = \{z \in \mathbb{R}^{|\boldsymbol{D}_S|} : \|z\| = 1\}$, and we write $A \sim St_{\boldsymbol{D}_S}(\alpha, \Gamma)$, if the flattened tensor A is distributed as $St_{|\boldsymbol{D}_S|}(\alpha, \Gamma)$, i.e. if the characteristic function of A is*

$$\varphi_A(t) = \mathbb{E}[e^{\mathrm{i}\,\boldsymbol{t} \otimes A}] = e^{-\int_{\mathbb{S}^{|\boldsymbol{D}_S|-1}} |\boldsymbol{t} \otimes \boldsymbol{s}|^\alpha \Gamma(\mathrm{d}\boldsymbol{s})}, \quad t \in \mathbb{R}^{\boldsymbol{D}_S}$$

where $\mathrm{d}\boldsymbol{s}$ is considered flattened.

SM: some useful inequalities

During the proofs we will use without any mention the following inequality:

Lemma 1. *For any real values $\alpha, z_1, \ldots z_n \geq 0$ there exists a constant $C = C(\alpha, n)$ such that*

$$(z_1 + \cdots + z_n)^\alpha \leq C(z_1^\alpha + \cdots + z_n^\alpha)$$

Proof. Let $Z = \max\{z_1, \ldots, z_n\}$. Thus we get

$$(z_1 + \cdots + z_n)^\alpha \leq (nZ)^\alpha = n^\alpha Z^\alpha \leq n^\alpha(z_1^\alpha + \cdots + z_n^\alpha)$$

In particular $C = C(\alpha, n) = n^\alpha$. $\qquad\qquad\square$

We give an important intuition of the reason why the proofs that will follow work. Intuitively we will provide an alternative proof of the strong law of large numbers for Stable random variables using the de Finetti theorem regarding the exchangeability of sequences of random variables. To this end we will have to require that the expected value of the stochastic process is finite. Using the above Lemma 1, from assumption (4) we get

$$|\phi(s)|^\alpha \leq (a + b|s|^\beta)^\alpha \leq 2^\alpha(a^\alpha + b^\alpha|s|^{\beta\alpha})$$

When s is α-stable distributed with any skewness, scale and shift parameters, then

$$\mathbb{E}[|\phi(s)|^\alpha] \leq 2^\alpha a^\alpha + 2^\alpha b^\alpha \mathbb{E}[|s|^{\beta\alpha}] < \infty$$

which is finite since $\beta < 1$ thus $\beta\alpha < \alpha$. Then, the assumption $\beta < 1$ is essential to guarantee the existence of the expected value of the stochastic process. However, this assumption also allows us to apply Jensen's inequality in the following sense: for any positive random variable s,

$$\mathbb{E}[s^\beta] \leq (\mathbb{E}[s])^\beta$$

We will use these inequalities repeatedly during the proofs.

SM: the patch operator

Referring to the definition (2) of DCNN we said that for each $l \in [L]$, $f_{\star p^{(l)}}^{(l-1)} \in \mathbb{R}^{C^{(l-1)} \times \mathbf{G}^{(l)}}$ is a patch extracted from $f^{(l-1)}$ at output position $p^{(l)} \in [\mathbf{P}^{(l)}]$ and is defined by $f_{\star p^{(l)}}^{(l-1)} = f_{(1:C^{(l-1)}, \star p^{(l)})}^{(l-1)}$. The patch operator is a map $\star : [\mathbf{P}^{(l)}] \to (\mathbb{Z}_1 \times \cdots \times \mathbb{Z}_{S^{(l-1)}})^{\mathbf{G}^{(l)}}$ that for each $p^{(l)}$, returns $\star p^{(l)}$ that is a $S^{(l-1)}$ tensor of dimension $\mathbf{G}^{(l)}$ containing some spatial indexes of $f^{(l-1)}$ depending on the moving window chosen in the structure (padding, stride etc). Refer to "SM: the patch operator" for a more detailed explanation.

More precisely, for each $g^{(l)} \in \mathbf{G}^{(l)}$, $\star p_{g^{(l)}}^{(l)} = (i_1, \ldots, i_{S^{(l-1)}}) \in \mathbb{Z}_1 \times \cdots \times \mathbb{Z}_{S^{(l-1)}}$ where we could have $i_{s^{(l-1)}} < 1$ or $i_{s^{(l-1)}} > P_{s^{(l-1)}}$ for some $s^{(l-1)} \in [S^{(l-1)}]$ because extraction could happens outside of $f^{(l-1)}$. Thus, two possibilities are allowed: 1) there exists $s^{(l-1)} \in [S^{(l-1)}]$ such that $i_{s^{(l-1)}} < 1$ or $i_{s^{(l-1)}} > P_{s^{(l-1)}}$. In that case $(f_{\star p^{(l)}}^{(l-1)})_{(c^{(l-1)}, g^{(l)})} = f_{(c^{(l-1)}, \star p_{g^{(l)}}^{(l)})}^{(l-1)} = 0$ for each $c^{(l-1)} \in [C^{(l-1)}]$; 2) $i_{s^{(l-1)}} \in [P_{s^{(l-1)}}]$ for all $s^{(l-1)} \in [S^{(l-1)}]$. In that case there exist $p^{(l-1)} \in [\mathbf{P}^{(l-1)}]$ such that $\star p_{g^{(l)}}^{(l)} = p^{(l-1)}$, then $(f_{\star p^{(l)}}^{(l-1)})_{(c^{(l-1)}, g^{(l)})} = f_{(c^{(l-1)}, p^{(l-1)})}^{(l-1)} = 0$ for each $c^{(l-1)} \in [C^{(l-1)}]$.

SM A

Here we compute the limit distribution of $f_{(c^{(l)}, :)}^{(l)(1:K)}$ as $C \to \infty$. First we prove the following two theorems (see the corresponding appendix for the proof):

Theorem 3. *[SM A.1]* $f_{(c^{(1)}, :)}^{(1)(1:K)} \sim St_{\mathbf{P}^{(1)} \times K}(\alpha, \Gamma^{(1)})$, *where*

$$\Gamma^{(1)} = \|\sigma_b \mathbb{1}_{(\mathbf{P}^{(1)} \times K)}\|^\alpha \Psi^{(1)}(\mathbb{1}_{(\mathbf{P}^{(1)} \times K)})$$
$$+ \sum_{(c^{(0)}, g^{(1)}) \in [C^{(0)} \times \mathbf{G}^{(1)}]} \|\sigma_\omega (x_\star^{(1:K)})_{(c^{(0)}, g^{(1)})}\|^\alpha \Psi^{(1)}\left((x_\star^{(1:K)})_{(c^{(0)}, g^{(1)})}\right)$$

where $(x_\star^{(1:K)})_{(c^{(0)}, g^{(1)})} = \left[(x_{\star p^{(1)}}^{(k)})_{(c^{(0)}, g^{(1)})}\right]_{\{(p^{(1)}, k) \in [\mathbf{P}^{(1)} \times K]\}}$.

Theorem 4. *[SM A.2] For* $l = 2, \ldots, L$, $f_{(c^{(l)}, :)}^{(l)(1:K)} | f_{(1:C, :)}^{(l-1)(1:K)} \sim St_{\mathbf{P}^{(l)} \times K}(\alpha, \Gamma_C^{(l)})$, *where*

$$\Gamma_C^{(l)} = \|\sigma_b \mathbb{1}_{(\mathbf{P}^{(l)} \times K)}\|^\alpha \Psi^{(l)}\left(\mathbb{1}_{(\mathbf{P}^{(l)} \times K)}\right)$$
$$+ \frac{1}{C} \sum_{(c^{(l-1)}, g^{(l)}) \in [C \times \mathbf{G}^{(l)}]} \|\sigma_\omega \phi(f_\star^{(l-1)(1:K)})_{(c^{(l-1)}, g^{(l)})}\|^\alpha \Psi^{(l)}\left(\phi(f_\star^{(l-1)(1:K)})_{(c^{(l-1)}, g^{(l)})}\right)$$

with $(f_\star^{(l-1)(1:K)})_{(c^{(l-1)}, g^{(l)})} = \left[(f_{\star p^{(l)}}^{(l-1)(k)})_{(c^{(l-1)}, g^{(l)})}\right]_{\{(p^{(l)}, k) \in [\mathbf{P}^{(l)} \times K]\}}$.

Now, for $l \in [L]$ we compute the limit distribution of $f_{(c^{(l)}, :)}^{(l)(1:K)}$ as $C \to \infty$. First recall that for $l = 1$, being $f_{(c^{(1)}, :)}^{(1)(1:K)}$ referred only to the $C^{(0)}$ channels of the input layer, then $f_{(c^{(1)}, :)}^{(1)(1:K)} \xrightarrow{d}$

7

$\mathrm{St}_{\mathbf{P}^{(1)} \times K}(\alpha, \Gamma_\infty^{(1)})$ constantly as $C \to \infty$, where we have defined $\Gamma_\infty^{(1)} = \Gamma^{(1)}$. Thus we compute the limit for all the others layers. We prove that for each $l = 2, \ldots, L$, $f_{(c^{(l)},:)}^{(l)(1:K)} \xrightarrow{d} \mathrm{St}_{\mathbf{P}^{(l)} \times K}(\alpha, \Gamma_\infty^{(l)})$ as $C \to \infty$, where

$$\Gamma_\infty^{(l)} = \|\sigma_b \mathbb{1}_{(\mathbf{P}^{(l)} \times K)}\|^\alpha \Psi^{(l)}\Big(\mathbb{1}_{(\mathbf{P}^{(l)} \times K)}\Big) +$$
$$+ \int \sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} \|\sigma_\omega \phi(f_{g^{(l)}})\|^\alpha \Psi^{(l)}\Big(\phi(f_{g^{(l)}})\Big) q^{(l-1)}(\mathrm{d}\, f_{\{g^{(l)} \in [\mathbf{G}^{(l)}]\}})$$

with $f_{g^{(l)}} \in \mathbb{R}^{\mathbf{P}^{(l)} \times K}$ for each $g^{(l)} \in \mathbf{G}^{(l)}$ and $q^{(l-1)} = \mathrm{St}_{\mathbf{P}^{(l-1)} \times K}(\alpha, \Gamma_\infty^{(l-1)})$.

For the proof we will need the following proposition which is a direct consequence of Exercise 2.3.4 of Samoradnitsky and Taqqu [1994].

Proposition 1. *If $A \sim St_D(\alpha, \Gamma)$ then for each $\boldsymbol{u} \in \mathbb{R}^D$ the 1-dimensional r.v. $\langle \boldsymbol{u}, A \rangle \sim St(\alpha, \tau(\boldsymbol{u}), \sigma(\boldsymbol{u}), \mu(\boldsymbol{u}))$ where*

$$\sigma(\boldsymbol{u}) = \Big(\int_{\mathbb{S}^{D-1}} |\langle \boldsymbol{u}, \boldsymbol{s} \rangle|^\alpha \Gamma(\mathrm{d}\boldsymbol{s})\Big)^{1/\alpha}$$

$$\tau(\boldsymbol{u}) = \sigma(\boldsymbol{u})^{-1} \int_{\mathbb{S}^{D-1}} |\langle \boldsymbol{u}, \boldsymbol{s} \rangle|^\alpha sign(\langle \boldsymbol{u}, \boldsymbol{s} \rangle) \Gamma(\mathrm{d}\boldsymbol{s})$$

$$\mu(\boldsymbol{u}) = \begin{cases} 0 & \alpha \neq 1 \\ -\frac{2}{\pi} \int_{\mathbb{S}^{D-1}} \langle \boldsymbol{u}, \boldsymbol{s} \rangle \log(|\langle \boldsymbol{u}, \boldsymbol{s} \rangle|) \Gamma(\mathrm{d}\boldsymbol{s}) & \alpha = 1 \end{cases}$$

Proof. Fix $l = 2, \ldots, L$ and, for each C, let $h_C^{(l)}$ denote the de Finetti random probability measure of the exchangeable sequence $((f_\star^{(l)(1:K)})_{(c^{(l)},:)})_{c^{(l)} \geq 1}$, i.e. $(f_\star^{(l)(1:K)})_{(c^{(l)},:)} | h_C^{(l)} \overset{iid}{\sim} h_C^{(l)}$. Consider the induction hypothesis that as $C \to \infty$

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

where $q^{(l-1)} = \mathrm{St}_{\mathbf{P}^{(l-1)} \times K}(\alpha, \Gamma_\infty^{(l-1)})$ and the finite measure $\Gamma_\infty^{(l-1)}$ will be specified. For $l > 1$ and any $\mathbf{t}^{(l)} := [t_{p^{(l)}}^{(l)(k)}]_{\{(p^{(l)},k) \in [\mathbf{P}^{(l)} \times K]\}} \in \mathbb{R}^{\mathbf{P}^{(l)} \times K}$,

$$\varphi_{\big(f_{(c^{(l)},:)}^{(l)(1:K)}\big)}(\mathbf{t}^{(l)})$$
$$= \mathbb{E}\Big[\exp\Big\{i\mathbf{t}^{(l)} \otimes f_{(c^{(l)},:)}^{(l)(1:K)}\Big\}\Big]$$
$$= \mathbb{E}\Big[\mathbb{E}\Big[\exp\Big\{i\mathbf{t}^{(l)} \otimes f_{(c^{(l)},:)}^{(l)(1:K)}\Big\}\Big|f_{(1:C,:)}^{(l-1)(1:K)}\Big]\Big]$$
$$= \mathbb{E}\Big[\exp\Big\{-\int_{\mathbb{S}^{|\mathbf{P}^{(l)} \times K|-1}} |\mathbf{t}^{(l)} \otimes s^{(l)}|^\alpha \Gamma_C^{(l)}(\mathrm{d}\, s^{(l)})\Big\}\Big]$$
$$= \exp\Big\{-\sigma_b^\alpha |\mathbf{t}^{(l)} \otimes \mathbb{1}_{(\mathbf{P}^{(l)} \times K)}|^\alpha\Big\} \times$$
$$\times \mathbb{E}\Big[\exp\Big\{-\frac{\sigma_\omega^\alpha}{C} \sum_{(c^{(l-1)},g^{(l)}) \in [C \times \mathbf{G}^{(l)}]} |\mathbf{t}^{(l)} \otimes \phi(f_\star^{(l-1)(1:K)})_{(c^{(l-1)},g^{(l)})}|^\alpha\Big\}\Big] \quad\quad (6)$$
$$= \exp\Big\{-\sigma_b^\alpha |\mathbf{t}^{(l)} \otimes \mathbb{1}_{(\mathbf{P}^{(l)} \times K)}|^\alpha\Big\} \times$$
$$\times \mathbb{E}\Big[\mathbb{E}\Big[\exp\Big\{-\frac{\sigma_\omega^\alpha}{C} \sum_{(c^{(l-1)},g^{(l)}) \in [C \times \mathbf{G}^{(l)}]} |\mathbf{t}^{(l)} \otimes \phi(f_\star^{(l-1)(1:K)})_{(c^{(l-1)},g^{(l)})}|^\alpha\Big\}\Big|h_C^{(l-1)}\Big]\Big]$$
$$= \exp\Big\{-\sigma_b^\alpha |\mathbf{t}^{(l)} \otimes \mathbb{1}_{(\mathbf{P}^{(l)} \times K)}|^\alpha\Big\} \times$$
$$\times \mathbb{E}\Big[\Big(\int \exp\Big\{-\frac{\sigma_\omega^\alpha}{C} \sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} |\mathbf{t}^{(l)} \otimes \phi(f_{g^{(l)}})|^\alpha\Big\} h_C^{(l-1)}(\mathrm{d}\, f_{\{g^{(l)} \in [\mathbf{G}^{(l)}]\}})\Big)^C\Big]$$

where $f_{g^{(l)}} \in \mathbb{R}^{\mathbf{P}^{(l)} \times K}$ for each $g^{(l)} \in \mathbf{G}^{(l)}$. Hereafter we show the limiting behaviour. In order to do this we need the following lemmas:

L1) For each $l = 2, \ldots, L$, $\sup_C \int \sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} \|\phi(f_{g^{(l)}})\|^{\alpha} h_C^{(l-1)}(\mathrm{d} f_{g^{(l)} \in [\mathbf{G}^{(l)}]}) < \infty$

L1.1) There exists $\epsilon > 0$ such that, $\sup_C \mathbb{E}[\sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} \|\phi(f_{\star}^{(l-1)(1:K)})_{(c^{(l-1)}, g^{(l)})}\|^{\alpha + \epsilon} |h_C^{(l-2)}] < \infty$ for each $l = 2, \ldots, L$

L2) $\int \sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} |\mathbf{t}^{(l)} \otimes \phi(f_{g^{(l)}})|^{\alpha} h_C^{(l-1)}(\mathrm{d} f_{g^{(l)} \in [\mathbf{G}^{(l)}]}) \xrightarrow{p} \int \sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} |\mathbf{t}^{(l)} \otimes \phi(f_{g^{(l)}})|^{\alpha} q^{(l-1)}(\mathrm{d} f_{g^{(l)} \in [\mathbf{G}^{(l)}]})$ as $C \to \infty$

L3) $\int \sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} \|\phi(f_{g^{(l)}})\|^{\alpha} [1 - \exp\{-\frac{\sigma_\omega^\alpha}{C} \sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} |\mathbf{t}^{(l)} \otimes \phi(f_{g^{(l)}})|^{\alpha}\}] h_C^{(l-1)}(\mathrm{d} f_{g^{(l)} \in [\mathbf{G}^{(l)}]}) \xrightarrow{p} 0$ as $C \to \infty$

Proof of L1

For $l = 2$, for each $c^{(1)} \geq 1$, from assumptions (3) and (4) and from Lemma 1 we get

$$
\mathbb{E}\Big[\sum_{g^{(2)} \in [\mathbf{G}^{(2)}]} \|\phi(f_\star^{(1)(1:K)})_{(c^{(1)}, g^{(2)})}\|^{\alpha}\Big]
$$

$$
= \mathbb{E}\Big[\sum_{g^{(2)} \in [\mathbf{G}^{(2)}]} \sum_{p^{(2)} \in [\mathbf{P}^{(2)}]} \sum_{k \in [K]} |\phi(f_{\star p^{(2)}}^{(1)(k)})_{(c^{(1)}, g^{(2)})}|^{\alpha}\Big]
$$

$$
= \sum_{g^{(2)} \in [\mathbf{G}^{(2)}]} \sum_{p^{(2)} \in [\mathbf{P}^{(2)}]} \sum_{k \in [K]} \mathbb{E}\Big[|\phi(f_{\star p^{(2)}}^{(1)(k)})_{(c^{(1)}, g^{(2)})}|^{\alpha}\Big]
$$

$$
\leq \sum_{g^{(2)} \in [\mathbf{G}^{(2)}]} \sum_{p^{(2)} \in [\mathbf{P}^{(2)}]} \sum_{k \in [K]} \mathbb{E}\Big[(a + b|(f_{\star p^{(2)}}^{(1)(k)})_{(c^{(1)}, g^{(2)})}|^{\beta})^{\alpha}\Big]
$$

$$
\leq 2^{\alpha} \sum_{g^{(2)} \in [\mathbf{G}^{(2)}]} \sum_{p^{(2)} \in [\mathbf{P}^{(2)}]} \sum_{k \in [K]} \mathbb{E}\Big[a^{\alpha} + b^{\alpha}|(f_{\star p^{(2)}}^{(1)(k)})_{(c^{(1)}, g^{(2)})}|^{\alpha\beta}\Big]
$$

$$
= 2^{\alpha} |\mathbf{G}^{(2)}| |\mathbf{P}^{(2)}| K a^{\alpha} + 2^{\alpha} b^{\alpha} \sum_{g^{(2)} \in [\mathbf{G}^{(2)}]} \sum_{p^{(2)} \in [\mathbf{P}^{(2)}]} \sum_{k \in [K]} \mathbb{E}\Big[|(f_{\star p^{(2)}}^{(1)(k)})_{(c^{(1)}, g^{(2)})}|^{\alpha\beta}\Big]
$$

$$
< \infty
$$

where we used that $(f_{\star p^{(2)}}^{(1)(k)})_{(c^{(1)}, g^{(2)})}$, by Proposition 1 is distributed according to a SD with index α (and some skewness, scale and shift parameters) and then, being $\alpha\beta < \alpha$, $\mathbb{E}\Big[|(f_{\star p^{(2)}}^{(1)(k)})_{(c^{(1)}, g^{(2)})}|^{\alpha\beta}\Big] < +\infty$. Now assuming that L1) is true for $l - 2$ we prove that it is true for $l - 1$. First, from assumptions (3) and (4) and from Lemma 1, we compute the following

$$
\mathbb{E}\Big[\sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} \|\phi(f_\star^{(l-1)(1:K)})_{(c^{(l-1)}, g^{(l)})}\|^{\alpha} \Big| f_{(1:C,:)}^{(l-2)(1:K)}\Big]
$$

$$
= \mathbb{E}\Big[\sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} \sum_{p^{(l)} \in [\mathbf{P}^{(l)}]} \sum_{k \in [K]} |\phi(f_{\star p^{(l)}}^{(l-1)(k)})_{(c^{(l-1)}, g^{(l)})}|^{\alpha} \Big| f_{(1:C,:)}^{(l-2)(1:K)}\Big]
$$

$$
= \sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} \sum_{p^{(l)} \in [\mathbf{P}^{(l)}]} \sum_{k \in [K]} \mathbb{E}\Big[|\phi(f_{\star p^{(l)}}^{(l-1)(k)})_{(c^{(l-1)}, g^{(l)})}|^{\alpha} \Big| f_{(1:C,:)}^{(l-2)(1:K)}\Big]
$$

$$
\leq \sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} \sum_{p^{(l)} \in [\mathbf{P}^{(l)}]} \sum_{k \in [K]} \mathbb{E}\Big[(a + b|(f_{\star p^{(l)}}^{(l-1)(k)})_{(c^{(l-1)}, g^{(l)})}|^{\beta})^{\alpha} \Big| f_{(1:C,:)}^{(l-2)(1:K)}\Big]
$$

$$
\leq 2^{\alpha} |\mathbf{G}^{(l)}| |\mathbf{P}^{(l)}| K a^{\alpha} + 2^{\alpha} b^{\alpha} \sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} \sum_{p^{(l)} \in [\mathbf{P}^{(l)}]} \sum_{k \in [K]} \mathbb{E}\Big[|(f_{\star p^{(l)}}^{(l-1)(k)})_{(c^{(l-1)}, g^{(l)})}|^{\beta\alpha} \Big| f_{(1:C,:)}^{(l-2)(1:K)}\Big]
$$

9

Recall that for each $p^{(l)} \in [\mathbf{P}^{(l)}]$, $(f_{\star p^{(l)}}^{(l-1)(k)})_{(c^{(l-1)}, g^{(l)})}$ could be equal to 0 or there exists an unique position $p^{(l-1)} \in [\mathbf{P}^{(l-1)}]$ such that $(f_{\star p^{(l)}}^{(l-1)(k)})_{(c^{(l-1)}, g^{(l)})} = f_{(c^{(l-1)}, p^{(l-1)})}^{(l-1)(k)}$, thus we get

$$
\mathbb{E}\Big[\sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} \|\phi(f_\star^{(l-1)(1:K)})_{(c^{(l-1)}, g^{(l)})}\|^\alpha \Big| f_{(1:C,:)}^{(l-2)(1:K)} \Big]
$$

$$
\leq 2^\alpha |\mathbf{G}^{(l)}||\mathbf{P}^{(l)}| K a^\alpha + 2^\alpha b^\alpha \sum_{p^{(l-1)} \in [\mathbf{P}^{(l-1)}]} \sum_{k \in [K]} \mathbb{E}\Big[|f_{(c^{(l-1)}, p^{(l-1)})}^{(l-1)(k)}|^{\beta\alpha} \Big| f_{(1:C,:)}^{(l-2)(1:K)} \Big]
$$

Moreover, from theorem 4 we know that $f_{(c^{(l-1)},:)}^{(l-1)(1:K)} | f_{(1:C,:)}^{(l-2)(1:K)} \sim \mathrm{St}_{\mathbf{P}^{(l-1)} \times K}(\alpha, \Gamma^{(l-1)})$ and from proposition 1, denoted $U(p^{(l-1)}, k) = \mathbb{1}_{(\mathbf{P}^{(l-1)} \times K)}[(p^{(l-1)}, k)]$, for each $(p^{(l-1)}, k) \in [\mathbf{P}^{(l-1)} \times K]$ we have $f_{(c^{(l-1)}, p^{(l-1)})}^{(l-1)(k)} | f_{(1:C,:)}^{(l-2)(1:K)} = U(p^{(l-1)}, k) \otimes f_{(c^{(l-1)},:)}^{(l-1)(1:K)} | f_{(1:C,:)}^{(l-2)(1:K)} \sim \mathrm{St}\Big(\alpha, \tau(U(p^{(l-1)}, k)), \sigma(U(p^{(l-1)}, k)), \mu(U(p^{(l-1)}, k))\Big) \sim \sigma(U(p^{(l-1)}, k)) \mathrm{St}\Big(\alpha, \tau(U(p^{(l-1)}, k)), 1, \mu(U(p^{(l-1)}, k))\Big)$. Since $\beta\alpha < \alpha$ we get

$$
\mathbb{E}\Big[\sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} \|\phi(f_\star^{(l-1)(1:K)})_{(c^{(l-1)}, g^{(l)})}\|^\alpha \Big| f_{(1:C,:)}^{(l-2)(1:K)} \Big]
$$

$$
\leq 2^\alpha |\mathbf{G}^{(l)}||\mathbf{P}^{(l)}| K a^\alpha +
$$

$$
+ 2^\alpha b^\alpha \sum_{p^{(l-1)} \in [\mathbf{P}^{(l-1)}]} \sum_{k \in [K]} \sigma\Big(U(p^{(l-1)}, k)\Big)^{\beta\alpha} \underbrace{\mathbb{E}\Big[|\mathrm{St}(\alpha, \tau(U(p^{(l-1)}, k)), 1, \mu(U(p^{(l-1)}, k)))|^{\beta\alpha} \Big]}_{<\infty}
$$

$$
\leq 2^\alpha |\mathbf{G}^{(l)}||\mathbf{P}^{(l)}| K a^\alpha +
$$

$$
+ 2^\alpha b^\alpha \mathcal{M} \sum_{p^{(l-1)} \in [\mathbf{P}^{(l-1)}]} \sum_{k \in [K]} \Big(\int_{\mathbb{S}^{|\mathbf{P}^{(l-1)} \times K| - 1}} \big| U(p^{(l-1)}, k) \otimes s \big|^\alpha \Gamma^{(l-1)}(\mathrm{d}\,s) \Big)^\beta
$$

where $\mathcal{M} = \max_{(p^{(l-1)}, k) \in [\mathbf{P}^{(l-1)} \times k]} \mathbb{E}\Big[|\mathrm{St}(\alpha, \tau(U(p^{(l-1)}, k)), 1, \mu(U(p^{(l-1)}, k)))|^{\beta\alpha} \Big] < +\infty$. Then,

$$
\mathbb{E}\Big[\sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} \|\phi(f_\star^{(l-1)(1:K)})_{(c^{(l-1)}, g^{(l)})}\|^\alpha \Big| h_C^{(l-2)} \Big]
$$

$$
= \mathbb{E}\Big[\mathbb{E}\Big[\sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} \|\phi(f_\star^{(l-1)(1:K)})_{(c^{(l-1)}, g^{(l)})}\|^\alpha \Big| f_{(1:C,:)}^{(l-2)(1:K)} \Big] \Big| h_C^{(l-2)} \Big]
$$

$$
\leq 2^\alpha |\mathbf{G}^{(l)}||\mathbf{P}^{(l)}| K a^\alpha +
$$

$$
+ 2^\alpha b^\alpha \mathcal{M} \sum_{p^{(l-1)} \in [\mathbf{P}^{(l-1)}]} \sum_{k \in [K]} \mathbb{E}\Big[\Big(\int_{\mathbb{S}^{|\mathbf{P}^{(l-1)} \times K| - 1}} \big| U(p^{(l-1)}, k) \otimes s \big|^\alpha \Gamma^{(l-1)}(\mathrm{d}\,s) \Big)^\beta \Big| h_C^{(l-2)} \Big]
$$

$$
\leq 2^\alpha |\mathbf{G}^{(l)}||\mathbf{P}^{(l)}| K a^\alpha +
$$

$$
+ 2^\alpha b^\alpha \mathcal{M} \sum_{p^{(l-1)} \in [\mathbf{P}^{(l-1)}]} \sum_{k \in [K]} \Big(\mathbb{E}\Big[\int_{\mathbb{S}^{|\mathbf{P}^{(l-1)} \times K| - 1}} \big| U(p^{(l-1)}, k) \otimes s \big|^\alpha \Gamma^{(l-1)}(\mathrm{d}\,s) \Big| h_C^{(l-2)} \Big] \Big)^\beta
$$

$$
\tag{7}
$$

where we used the Jensen's inequality. Moreover,

$$\mathbb{E}\left[\int_{\mathbb{S}^{|\mathbf{P}^{(l-1)}\times K|-1}}\left|U(p^{(l-1)},k)\otimes s\right|^{\alpha}\Gamma^{(l-1)}(\mathrm{d}\,s)\bigg|h_C^{(l-2)}\right]$$

$$=\mathbb{E}\left[\sigma_b^{\alpha}|U(p^{(l-1)},k)\otimes\mathbb{1}_{(\mathbf{P}^{(l-1)}\times K)}|^{\alpha}+\right.$$

$$\left.+\frac{\sigma_{\omega}^{\alpha}}{C}\sum_{(c^{(l-2)},g^{(l-1)})\in[C\times\mathbf{G}^{(l-1)}]}\left|U(p^{(l-1)},k)\otimes\phi(f_{\star}^{(l-2)(1:K)})_{(c^{(l-2)},g^{(l-1)})}\right|^{\alpha}\bigg|h_C^{(l-2)}\right]$$

$$=\mathbb{E}\left[\sigma_b^{\alpha}+\frac{\sigma_{\omega}^{\alpha}}{C}\sum_{(c^{(l-2)},g^{(l-1)})\in[C\times\mathbf{G}^{(l-1)}]}|\phi(f_{\star p^{(l-1)}}^{(l-2)(k)})_{(c^{(l-2)},g^{(l-1)})}|^{\alpha}\bigg|h_C^{(l-2)}\right] \tag{8}$$

$$=\sigma_b^{\alpha}+\frac{\sigma_{\omega}^{\alpha}}{C}\sum_{(c^{(l-2)},g^{(l-1)})\in[C\times\mathbf{G}^{(l-1)}]}\mathbb{E}\left[|\phi(f_{\star p^{(l-1)}}^{(l-2)(k)})_{(c^{(l-2)},g^{(l-1)})}|^{\alpha}\bigg|h_C^{(l-2)}\right]$$

$$=\sigma_b^{\alpha}+\frac{\sigma_{\omega}^{\alpha}}{C}\sum_{(c^{(l-2)},g^{(l-1)})\in[C\times\mathbf{G}^{(l-1)}]}\mathbb{E}\left[\|\phi(f_{\star}^{(l-2)(1:K)})_{(c^{(l-2)},g^{(l-1)})}\|^{\alpha}\bigg|h_C^{(l-2)}\right]$$

$$=\sigma_b^{\alpha}+\frac{\sigma_{\omega}^{\alpha}}{C}\sum_{(c^{(l-2)},g^{(l-1)})\in[C\times\mathbf{G}^{(l-1)}]}\int\|\phi(f_{g^{(l-1)}})\|^{\alpha}h_C^{(l-2)}(\mathrm{d}\,f_{g^{(l-1)}\in[\mathbf{G}^{(l-1)}]})$$

$$=\sigma_b^{\alpha}+\sigma_{\omega}^{\alpha}\int\sum_{g^{(l-1)}\in[\mathbf{G}^{(l-1)}]}\|\phi(f_{g^{(l-1)}})\|^{\alpha}h_C^{(l-2)}(\mathrm{d}\,f_{g^{(l-1)}\in[\mathbf{G}^{(l-1)}]})$$

Note that we have used the inequality $|x_i|\leq\|x\|$ and that $(f_{\star}^{(l-2)(1:K)})_{(c^{(l-2)},:)}|h_C^{(l-2)}\overset{iid}{\sim}h_C^{(l-2)}$ (with respect to $c^{(l-1)}\geq 1$). Putting together (7) and (8) we have shown that

$$\sup_C\mathbb{E}\left[\sum_{g^{(l)}\in[\mathbf{G}^{(l)}]}\|\phi(f_{\star}^{(l-1)(1:K)})_{(c^{(l-1)},g^{(l)})}\|^{\alpha}\bigg|h_C^{(l-2)}\right]$$

$$\leq 2^{\alpha}|\mathbf{G}^{(l)}||\mathbf{P}^{(l)}|Ka^{\alpha}+2^{\alpha}b^{\alpha}\mathcal{M}\sum_{p^{(l-1)}\in[\mathbf{P}^{(l-1)}]}\sum_{k\in[K]}\left(\sigma_b^{\alpha}\right.$$

$$\left.+\sigma_{\omega}^{\alpha}\sup_C\int\sum_{g^{(l-1)}\in[\mathbf{G}^{(l-1)}]}\|\phi(f_{g^{(l-1)}})\|^{\alpha}h_C^{(l-2)}(\mathrm{d}\,f_{g^{(l-1)}\in[\mathbf{G}^{(l-1)}]})\right)^{\beta}$$

which is finite by induction hypothesis. Now we conclude:

$$\sup_C\int\sum_{g^{(l)}\in[\mathbf{G}^{(l)}]}\|\phi(f_{g^{(l)}})\|^{\alpha}h_C^{(l-1)}(\mathrm{d}\,f_{g^{(l)}\in[\mathbf{G}^{(l)}]})$$

$$=\sup_C\mathbb{E}\left[\sum_{g^{(l)}\in[\mathbf{G}^{(l)}]}\|\phi(f_{\star}^{(l-1)(1:K)})_{(c^{(l-1)},g^{(l)})}\|^{\alpha}\bigg|h_C^{(l-1)}\right]$$

$$=\sup_C\mathbb{E}\left[\sup_C\mathbb{E}\left[\sum_{g^{(l)}\in[\mathbf{G}^{(l)}]}\|\phi(f_{\star}^{(l-1)(1:K)})_{(c^{(l-1)},g^{(l)})}\|^{\alpha}\bigg|h_C^{(l-2)}\right]\bigg|h_C^{(l-1)}\right]$$

$$<+\infty$$

that is finite by previous step.

Proof of L1.1

The proof of L1.1) follows by induction, and along lines similar to the proof of L1). In particular, let $\epsilon > 0$ be such that $\beta(\alpha + \epsilon) < \alpha$. It exists since $\beta < 1$. For $l = 2$, for each $c^{(1)} \geq 1$

$$\mathbb{E}\Big[\sum_{g^{(2)} \in [\mathbf{G}^{(2)}]} \|\phi(f_\star^{(1)(1:K)})_{(c^{(1)}, g^{(2)})}\|^{\alpha + \epsilon} \Big]$$

$$= \mathbb{E}\Big[\sum_{g^{(2)} \in [\mathbf{G}^{(2)}]} \sum_{p^{(2)} \in [\mathbf{P}^{(2)}]} \sum_{k \in [K]} |\phi(f_{\star p^{(2)}}^{(1)(k)})_{(c^{(1)}, g^{(2)})}|^{\alpha + \epsilon} \Big]$$

$$= \sum_{g^{(2)} \in [\mathbf{G}^{(2)}]} \sum_{p^{(2)} \in [\mathbf{P}^{(2)}]} \sum_{k \in [K]} \mathbb{E}\Big[|\phi(f_{\star p^{(2)}}^{(1)(k)})_{(c^{(1)}, g^{(2)})}|^{\alpha + \epsilon} \Big]$$

$$\leq \sum_{g^{(2)} \in [\mathbf{G}^{(2)}]} \sum_{p^{(2)} \in [\mathbf{P}^{(2)}]} \sum_{k \in [K]} \mathbb{E}\Big[(a + b|(f_{\star p^{(2)}}^{(1)(k)})_{(c^{(1)}, g^{(2)})}|^\beta)^{(\alpha + \epsilon)} \Big]$$

$$\leq \sum_{g^{(2)} \in [\mathbf{G}^{(2)}]} \sum_{p^{(2)} \in [\mathbf{P}^{(2)}]} \sum_{k \in [K]} \mathbb{E}\Big[a^{(\alpha + \epsilon)} + b^{(\alpha + \epsilon)}|(f_{\star p^{(2)}}^{(1)(k)})_{(c^{(1)}, g^{(2)})}|^{(\alpha + \epsilon)\beta} \Big]$$

$$\leq 2^{(\alpha + \epsilon)}|\mathbf{G}^{(2)}||\mathbf{P}^{(2)}|Ka^{(\alpha + \epsilon)} + (2b)^{(\alpha + \epsilon)} \sum_{g^{(2)} \in [\mathbf{G}^{(2)}]} \sum_{p^{(2)} \in [\mathbf{P}^{(2)}]} \sum_{k \in [K]} \mathbb{E}\Big[|(f_{\star p^{(2)}}^{(1)(k)})_{(c^{(1)}, g^{(2)})}|^{(\alpha + \epsilon)\beta} \Big]$$

$$< \infty$$

where we used that $(f_{\star p^{(2)}}^{(1)(k)})_{(c^{(1)}, g^{(2)})}$, by Proposition 1 is distributed according to a SD with index α (and some skewness, scale and shift parameters) and then, being $(\alpha + \epsilon)\beta < \alpha$, $\mathbb{E}\Big[|(f_{\star p^{(2)}}^{(1)(k)})_{(c^{(1)}, g^{(2)})}|^{(\alpha + \epsilon)\beta} \Big] < +\infty$. Moreover the bound is uniform with respect to C since the law is invariant with respect to C. Now assuming that L1.1) is true for $l - 2$ we prove that it is true for $l - 1$. As in the previous lemma, we can write the following inequality

$$\mathbb{E}\Big[\sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} \|\phi(f_\star^{(l-1)(1:K)})_{(c^{(l-1)}, g^{(l)})}\|^{\alpha + \epsilon} \Big| h_C^{(l-2)} \Big]$$

$$= \mathbb{E}\Big[\mathbb{E}\Big[\sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} \|\phi(f_\star^{(l-1)(1:K)})_{(c^{(l-1)}, g^{(l)})}\|^{\alpha + \epsilon} \Big| f_{(1:C,:)}^{(l-2)(1:K)} \Big] \Big| h_C^{(l-2)} \Big]$$

$$\leq 2^{(\alpha + \epsilon)}|\mathbf{G}^{(l)}||\mathbf{P}^{(l)}|Ka^{(\alpha + \epsilon)} + 2^{(\alpha + \epsilon)}b^{(\alpha + \epsilon)}\mathcal{M} \sum_{p^{(l-1)} \in [\mathbf{P}^{(l-1)}]}$$

$$\sum_{k \in [K]} \mathbb{E}\Big[\Big(\int_{\mathbb{S}^{|\mathbf{P}^{(l-1)}| \times K| - 1}} |U(p^{(l-1)}, k) \otimes s|^{\alpha + \epsilon} \Gamma^{(l-1)}(\mathrm{d}\,s) \Big)^\beta \Big| h_C^{(l-2)} \Big] \qquad (9)$$

$$\leq 2^{(\alpha + \epsilon)}|\mathbf{G}^{(l)}||\mathbf{P}^{(l)}|Ka^{(\alpha + \epsilon)} + 2^{(\alpha + \epsilon)}b^{(\alpha + \epsilon)}\mathcal{M} \sum_{p^{(l-1)} \in [\mathbf{P}^{(l-1)}]}$$

$$\sum_{k \in [K]} \Big(\mathbb{E}\Big[\int_{\mathbb{S}^{|\mathbf{P}^{(l-1)}| \times K| - 1}} |U(p^{(l-1)}, k) \otimes s|^{\alpha + \epsilon} \Gamma^{(l-1)}(\mathrm{d}\,s) \Big| h_C^{(l-2)} \Big] \Big)^\beta$$

Moreover, following the same steps as in the previous lemma (just replacing $\alpha + \epsilon$ instead of α), we get

$$\mathbb{E}\Big[\int_{\mathbb{S}^{|\mathbf{P}^{(l-1)}| \times K| - 1}} |U(p^{(l-1)}, k) \otimes s|^{\alpha + \epsilon} \Gamma^{(l-1)}(\mathrm{d}\,s) \Big| h_C^{(l-2)} \Big]$$

$$= \sigma_b^{\alpha + \epsilon} + \sigma_\omega^{\alpha + \epsilon} \int \sum_{g^{(l-1)} \in [\mathbf{G}^{(l-1)}]} \|\phi(f_{g^{(l-1)}})\|^{\alpha + \epsilon} h_C^{(l-2)}(\mathrm{d}\,f_{g^{(l-1)} \in [\mathbf{G}^{(l-1)}]})$$

and

$$\int \sum_{g^{(l-1)}\in[\mathbf{G}^{(l-1)}]} \|\phi(f_{g^{(l-1)}})\|^{\alpha+\epsilon} h_C^{(l-2)}(\mathrm{d}\,f_{g^{(l-1)}\in[\mathbf{G}^{(l-1)}]})$$

$$= \mathbb{E}\Big[\sum_{g^{(l-1)}\in[\mathbf{G}^{(l-1)}]} \|\phi(f_\star^{(l-2)(1:K)})_{(c^{(l-2)},g^{(l-1)})}\|^{\alpha+\epsilon}\Big|h_C^{(l-2)}\Big]$$

$$= \mathbb{E}\Big[\mathbb{E}\Big[\sum_{g^{(l-1)}\in[\mathbf{G}^{(l-1)}]} \|\phi(f_\star^{(l-2)(1:K)})_{(c^{(l-2)},g^{(l-1)})}\|^{\alpha+\epsilon}\Big|h_C^{(l-3)}\Big]\Big|h_C^{(l-2)}\Big]$$

$$\leq \mathbb{E}\Big[\mathbb{E}\Big[\sum_{g^{(l-1)}\in[\mathbf{G}^{(l-1)}]} \|\phi(f_\star^{(l-2)(1:K)})_{(c^{(l-2)},g^{(l-1)})}\|^{\alpha+\epsilon}\Big|h_C^{(l-3)}\Big]\Big|h_C^{(l-2)}\Big]$$

Thus taking the $\sup_C$ in (9), by previous inequalities, it is less or equal than

$$2^{(\alpha+\epsilon)}|\mathbf{G}^{(l)}||\mathbf{P}^{(l)}|Ka^{(\alpha+\epsilon)} + 2^{(\alpha+\epsilon)}b^{(\alpha+\epsilon)}\mathcal{M}\sum_{p^{(l-1)}\in[\mathbf{P}^{(l-1)}]}\sum_{k\in[K]}\Big(\sigma_b^{\alpha+\epsilon}+$$

$$+ \sigma_\omega^{\alpha+\epsilon}\mathbb{E}\Big[\sup_C \mathbb{E}\Big[\sum_{g^{(l-1)}\in[\mathbf{G}^{(l-1)}]} \|\phi(f_\star^{(l-2)(1:K)})_{(c^{(l-2)},g^{(l-1)})}\|^{\alpha+\epsilon}\Big|h_C^{(l-3)}\Big]\Big|h_C^{(l-2)}\Big]\Big)^\beta$$

which is bounded by hypothesis induction.

Proof of L2

By induction hypothesis, $h_C^{(l-1)}$ converges to $h^{(l-1)}$ in distribution with respect to the weak topology. Since the limit law is degenerate on $h^{(l-1)}$ (in the sense that it provides a.s. the distribution $q^{(l-1)}$), then for every sub-sequence (C') there exists a sub-sequence (C'') such that $h_{C''}^{(l-1)}$ converges a.s. By the induction hypothesis, $h^{(l-1)}$ is absolutely continuous with respect to the Lebesgue measure. Since ϕ is almost everywhere continuous, and by L1.1) uniformly integrable with respect to $(h_C^{(l-1)})$ then we can write the following

$$\int \sum_{g^{(l)}\in[\mathbf{G}^{(l)}]} |\mathbf{t}^{(l)}\otimes\phi(f_{g^{(l)}})|^\alpha h_{C''}^{(l-1)}(\mathrm{d}\,f_{g^{(l)}\in[\mathbf{G}^{(l)}]}) \xrightarrow{a.s.} \int \sum_{g^{(l)}\in[\mathbf{G}^{(l)}]} |\mathbf{t}^{(l)}\otimes\phi(f_{g^{(l)}})|^\alpha q^{(l-1)}(\mathrm{d}\,f_{g^{(l)}\in[\mathbf{G}^{(l)}]})$$

Thus

$$\int \sum_{g^{(l)}\in[\mathbf{G}^{(l)}]} |\mathbf{t}^{(l)}\otimes\phi(f_{g^{(l)}})|^\alpha h_C^{(l-1)}(\mathrm{d}\,f_{g^{(l)}\in[\mathbf{G}^{(l)}]}) \xrightarrow{p} \int \sum_{g^{(l)}\in[\mathbf{G}^{(l)}]} |\mathbf{t}^{(l)}\otimes\phi(f_{g^{(l)}})|^\alpha q^{(l-1)}(\mathrm{d}\,f_{g^{(l)}\in[\mathbf{G}^{(l)}]})$$

as $C \to +\infty$.

Proof of L3

Let $\epsilon > 0$ as in L1.1), $r = \frac{\alpha+\epsilon}{\alpha}$ and q such that $\frac{1}{r} + \frac{1}{q} = 1$. Thus, by Hölder inequality and by the fact that, being $q > 1$, for every $y \geq 0$ it holds that $(1-e^{-y})^q \leq (1-e^{-y}) \leq y$, we get

$$\int \sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} \|\phi(f_{g^{(l)}})\|^{\alpha} [1 - \exp\{-\frac{\sigma_\omega^\alpha}{C} \sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} |\mathbf{t}^{(l)} \otimes \phi(f_{g^{(l)}})|^{\alpha}\}] h_C^{(l-1)}(\mathrm{d}\, f_{g^{(l)} \in [\mathbf{G}^{(l)}]})$$

$$\leq \sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} \left[\int \|\phi(f_{g^{(l)}})\|^{\alpha r} h_C^{(l-1)}(\mathrm{d}\, f_{g^{(l)}}) \right]^{\frac{1}{r}} \times$$

$$\times \left[\int [1 - \exp\{-\frac{\sigma_\omega^\alpha}{C} \sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} |\mathbf{t}^{(l)} \otimes \phi(f_{g^{(l)}})|^{\alpha}\}]^q h_C^{(l-1)}(\mathrm{d}\, f_{g^{(l)}}) \right]^{\frac{1}{q}}$$

$$\leq \sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} \left[\int \|\phi(f_{g^{(l)}})\|^{\alpha r} h_C^{(l-1)}(\mathrm{d}\, f_{g^{(l)}}) \right]^{\frac{1}{r}} \left[\frac{\sigma_\omega^\alpha}{C} \int \sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} |\mathbf{t}^{(l)} \otimes \phi(f_{g^{(l)}})|^{\alpha} h_C^{(l-1)}(\mathrm{d}\, f_{g^{(l)}}) \right]^{\frac{1}{q}}$$

$$\leq \sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} \left[\int \|\phi(f_{g^{(l)}})\|^{\alpha r} h_C^{(l-1)}(\mathrm{d}\, f_{g^{(l)}}) \right]^{\frac{1}{r}} \times$$

$$\times \left[\|\mathbf{t}^{(l)}\|^{\alpha} \frac{\sigma_\omega^\alpha}{C} \int \sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} \|\phi(f_{g^{(l)}})\|^{\alpha} h_C^{(l-1)}(\mathrm{d}\, f_{g^{(l)}}) \right]^{\frac{1}{q}}$$

$$= \left(\|\mathbf{t}^{(l)}\|^{\alpha} \frac{\sigma_\omega^\alpha}{C} \right)^{\frac{1}{q}} \sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} \left[\int \|\phi(f_{g^{(l)}})\|^{\alpha+\epsilon} h_C^{(l-1)}(\mathrm{d}\, f_{g^{(l)}}) \right]^{\frac{1}{r}} \times$$

$$\times \left[\int \sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} \|\phi(f_{g^{(l)}})\|^{\alpha} h_C^{(l-1)}(\mathrm{d}\, f_{g^{(l)}}) \right]^{\frac{1}{q}} \xrightarrow{p} 0$$

as $C \to \infty$ by L1) and L1.1).

Conclusion: L1 + L1.1 + L2 + L3

By Lagrange theorem for $y > 0$ there exists $\theta \in (0,1)$ such that $e^{-y} = 1 - y + y(1 - e^{-y\theta})$. In our case, for $y = y_C(f_{g^{(l)} \in [\mathbf{G}^{(l)}]}) = \frac{\sigma_\omega^\alpha}{C} \sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} |\mathbf{t}^{(l)} \otimes \phi(f_{g^{(l)}})|^{\alpha}$, for any C there exists a $\theta_C \in (0,1)$ such that, from (6), the follow equality holds

$$\varphi_{\left(f_{(c^{(l)},:)}^{(l)(1:K)}\right)}(\mathbf{t}^{(l)})$$

$$=e^{-\sigma_b^\alpha |\mathbf{t}^{(l)} \otimes \mathbb{1}_{(\mathbf{P}^{(l)} \times K)}|^\alpha} \, \mathbb{E}\left[\left(\int e^{-y_C\left(f_{g^{(l)} \in [\mathbf{G}^{(l)}]}\right)} h_C^{(l-1)}(\mathrm{d}\, f_{\{g^{(l)} \in [\mathbf{G}^{(l)}]\}})\right)^C\right]$$

$$=e^{-\sigma_b^\alpha |\mathbf{t}^{(l)} \otimes \mathbb{1}_{(\mathbf{P}^{(l)} \times K)}|^\alpha} \, \mathbb{E}\left[\left(1 - \int y_C\left(f_{g^{(l)} \in [\mathbf{G}^{(l)}]}\right) h_C^{(l-1)}(\mathrm{d}\, f_{\{g^{(l)} \in [\mathbf{G}^{(l)}]\}})+\right.\right.$$

$$\left.\left. + \int y_C\left(f_{g^{(l)} \in [\mathbf{G}^{(l)}]}\right)\left[1 - e^{-\theta_C y_C\left(f_{g^{(l)} \in [\mathbf{G}^{(l)}]}\right)}\right] h_C^{(l-1)}(\mathrm{d}\, f_{\{g^{(l)} \in [\mathbf{G}^{(l)}]\}})\right)^C\right]$$

$$= \exp\left\{-\sigma_b^\alpha |\mathbf{t}^{(l)} \otimes \mathbb{1}_{(\mathbf{P}^{(l)} \times K)}|^\alpha\right\} \times$$

$$\times \mathbb{E}\left[\left(1 - \frac{\sigma_\omega^\alpha}{C}\int \sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} |\mathbf{t}^{(l)} \otimes \phi(f_{g^{(l)})}|^\alpha h_C^{(l-1)}(\mathrm{d}\, f_{\{g^{(l)} \in [\mathbf{G}^{(l)}]\}})+\right.\right.$$

$$+ \frac{\sigma_\omega^\alpha}{C}\int \sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} |\mathbf{t}^{(l)} \otimes \phi(f_{g^{(l)})}|^\alpha \left[1 - \exp\left\{-\theta_C \frac{\sigma_\omega^\alpha}{C}\sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} |\mathbf{t}^{(l)} \otimes \phi(f_{g^{(l)})}|^\alpha\right\}\right] \times$$

$$\left.\left. \times h_C^{(l-1)}(\mathrm{d}\, f_{\{g^{(l)} \in [\mathbf{G}^{(l)}]\}})\right)^C\right]$$

The last integral tends to 0 as $C \to \infty$ since by Cauchy inequality $|\mathbf{t}^{(l)} \otimes \phi(f_{g^{(l)}})|^\alpha \leq \|\mathbf{t}^{(l)}\|^\alpha \|\phi(f_{g^{(l)}})\|^\alpha$ we have

$$\int \sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} |\mathbf{t}^{(l)} \otimes \phi(f_{g^{(l)})}|^\alpha \left[1 - \exp\left\{-\theta_C \frac{\sigma_\omega^\alpha}{C}\sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} |\mathbf{t}^{(l)} \otimes \phi(f_{g^{(l)})}|^\alpha\right\}\right] \times$$

$$\times h_C^{(l-1)}(\mathrm{d}\, f_{\{g^{(l)} \in [\mathbf{G}^{(l)}]\}})$$

$$\leq \|\mathbf{t}^{(l)}\|^\alpha \int \sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} \|\phi(f_{g^{(l)}})\|^\alpha \left[1 - \exp\left\{-\theta_C \frac{\sigma_\omega^\alpha}{C}\sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} |\mathbf{t}^{(l)} \otimes \phi(f_{g^{(l)})}|^\alpha\right\}\right] \times$$

$$\times h_C^{(l-1)}(\mathrm{d}\, f_{\{g^{(l)} \in [\mathbf{G}^{(l)}]\}})$$

$$\leq \|\mathbf{t}^{(l)}\|^\alpha \int \sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} \|\phi(f_{g^{(l)}})\|^\alpha \left[1 - \exp\left\{-\frac{\sigma_\omega^\alpha}{C}\sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} |\mathbf{t}^{(l)} \otimes \phi(f_{g^{(l)})}|^\alpha\right\}\right] \times$$

$$\times h_C^{(l-1)}(\mathrm{d}\, f_{\{g^{(l)} \in [\mathbf{G}^{(l)}]\}})$$

which tends to 0 in probability by L3). Thus, by using the definition of the exponential function $e^x = \lim_{n\to\infty}(1 + x/n)^n$, and by L2), we get

$$\varphi_{\left(f_{(c^{(l)},:)}^{(l)(1:K)}\right)}(\mathbf{t}^{(l)})$$

$$\to \exp\left\{-\sigma_b^\alpha |\mathbf{t}^{(l)} \otimes \mathbb{1}_{(\mathbf{P}^{(l)} \times K)}|^\alpha - \sigma_\omega^\alpha \int \sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} |\mathbf{t}^{(l)} \otimes \phi(f_{g^{(l)})}|^\alpha q^{(l-1)}(\mathrm{d}\, f_{\{g^{(l)} \in [\mathbf{G}^{(l)}]\}})\right\}$$

$$= \exp\left\{-\sigma_b^\alpha \|\mathbb{1}_{(\mathbf{P}^{(l)} \times K)}\|^\alpha \left|\mathbf{t}^{(l)} \otimes \frac{\mathbb{1}_{(\mathbf{P}^{(l)} \times K)}}{\|\mathbb{1}_{(\mathbf{P}^{(l)} \times K)}\|}\right|^\alpha +\right.$$

$$\left. - \sigma_\omega^\alpha \int \sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} \|\phi(f_{g^{(l)}})\|^\alpha \left|\mathbf{t}^{(l)} \otimes \frac{\phi(f_{g^{(l)})}}{\|\phi(f_{g^{(l)})}\|}\right|^\alpha q^{(l-1)}(\mathrm{d}\, f_{\{g^{(l)} \in [\mathbf{G}^{(l)}]\}})\right\}$$

$$= \exp\left\{-\int_{\mathbb{S}^{|\mathbf{P}^{(l)} \times K|-1}} |\mathbf{t}^{(l)} \otimes s^{(l)}|^\alpha \Gamma_\infty^{(l)}(\mathrm{d}\, s^{(l)})\right\}$$

15

where

$$
\begin{aligned}
\Gamma^{(l)}_\infty =& \|\sigma_b \mathbb{1}_{(\mathbf{P}^{(l)} \times K)}\|^\alpha \Psi^{(l)}\Big(\mathbb{1}_{(\mathbf{P}^{(l)} \times K)} \Big)+ \\
&+ \int \sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} \|\sigma_\omega \phi(f_{g^{(l)}})\|^\alpha \Psi^{(l)}\Big(\phi(f_{g^{(l)}}) \Big) q^{(l-1)}(\mathrm{d}\, f_{\{g^{(l)} \in [\mathbf{G}^{(l)}]\}})
\end{aligned}
$$

$\square$

SM A.1

We prove that $f^{(1)(1:K)}_{(c^{(1)},:)} \sim \mathrm{St}_{\mathbf{P}^{(1)} \times K}(\alpha, \Gamma^{(1)})$, where

$$
\begin{aligned}
\Gamma^{(1)} =& \|\sigma_b \mathbb{1}_{(\mathbf{P}^{(1)} \times K)}\|^\alpha \Psi^{(1)}(\mathbb{1}_{(\mathbf{P}^{(1)} \times K)}) \\
&+ \sum_{(c^{(0)},g^{(1)}) \in [C^{(0)} \times \mathbf{G}^{(1)}]} \|\sigma_\omega (x^{(1:K)}_\star)_{(c^{(0)},g^{(1)})}\|^\alpha \Psi^{(1)}\Big((x^{(1:K)}_\star)_{(c^{(0)},g^{(1)})} \Big)
\end{aligned}
$$

Proof. For $l = 1$, from definition (5) and assumption (3) we have that, for any $\mathbf{t}^{(1)} := [t^{(1)(k)}_{p^{(1)}}]_{\{(p^{(1)},k) \in [\mathbf{P}^{(1)} \times K]\}} \in \mathbb{R}^{\mathbf{P}^{(1)} \times K}$

$$
\begin{aligned}
& \varphi_{(f^{(1)(1:K)}_{(c^{(1)},:)})}(\mathbf{t}^{(1)}) \\
&= \mathbb{E}\left[\exp\left\{ \mathrm{i}\, \mathbf{t}^{(1)} \otimes f^{(1)(1:K)}_{(c^{(1)},:)} \right\} \right] \\
&= \mathbb{E}\left[\exp\left\{ \mathrm{i}\, \mathbf{t}^{(1)} \otimes \Big(W^{(1)}_{(c^{(1)},:,:)} \overset{(\mathbf{P}^{(1)},K)}{\otimes} x^{(1:K)}_\star + b^{(1)}_{c^{(1)}} \mathbb{1}_{(\mathbf{P}^{(1)} \times K)} \Big) \right\} \right] \\
&= \mathbb{E}\left[\exp\left\{ \mathrm{i}\, W^{(1)}_{(c^{(1)},:,:)} \overset{(\mathbf{P}^{(1)},K)}{\otimes} \Big(\mathbf{t}^{(1)} \overset{(C^{(0)},\mathbf{G}^{(1)})}{\otimes} x^{(1:K)}_\star \Big) + b^{(1)}_{c^{(1)}} \mathbf{t}^{(1)} \otimes \mathbb{1}_{(\mathbf{P}^{(1)} \times K)} \right\} \right] \\
&= \mathbb{E}\left[\exp\left\{ b^{(1)}_{c^{(1)}} \mathbf{t}^{(1)} \otimes \mathbb{1}_{(\mathbf{P}^{(1)} \times K)} \right\} \right] \times \\
& \quad \times \prod_{(c^{(0)},g^{(1)}) \in [C^{(0)} \times \mathbf{G}^{(1)}]} \mathbb{E}\left[\exp\left\{ \mathrm{i}\, W^{(1)}_{(c^{(1)},c^{(0)},g^{(1)})} \mathbf{t}^{(1)} \otimes (x^{(1:K)}_\star)_{(c^{(0)},g^{(1)})} \right\} \right] \\
&= e^{-\sigma_b^\alpha |\mathbf{t}^{(1)} \otimes \mathbb{1}_{(\mathbf{P}^{(1)} \times K)}|^\alpha} \prod_{(c^{(0)},g^{(1)}) \in [C^{(0)} \times \mathbf{G}^{(1)}]} e^{-\sigma_\omega^\alpha \left| \mathbf{t}^{(1)} \otimes (x^{(1:K)}_\star)_{(c^{(0)},g^{(1)})} \right|^\alpha} \\
&= \exp\left\{ -\sigma_b^\alpha |\mathbf{t}^{(1)} \otimes \mathbb{1}_{(\mathbf{P}^{(1)} \times K)}|^\alpha - \sigma_\omega^\alpha \sum_{(c^{(0)},g^{(1)}) \in [C^{(0)} \times \mathbf{G}^{(1)}]} \left| \mathbf{t}^{(1)} \otimes (x^{(1:K)}_\star)_{(c^{(0)},g^{(1)})} \right|^\alpha \right\} \\
&= \exp\left\{ -\sigma_b^\alpha \| \mathbb{1}_{(\mathbf{P}^{(1)} \times K)}\|^\alpha \left| \mathbf{t}^{(1)} \otimes \frac{\mathbb{1}_{(\mathbf{P}^{(1)} \times K)}}{\| \mathbb{1}_{(\mathbf{P}^{(1)} \times K)}\|} \right|^\alpha + \right. \\
& \quad \left. - \sigma_\omega^\alpha \sum_{(c^{(0)},g^{(1)}) \in [C^{(0)} \times \mathbf{G}^{(1)}]} \|(x^{(1:K)}_\star)_{(c^{(0)},g^{(1)})}\|^\alpha \left| \mathbf{t}^{(1)} \otimes \frac{(x^{(1:K)}_\star)_{(c^{(0)},g^{(1)})}}{\|(x^{(1:K)}_\star)_{(c^{(0)},g^{(1)})}\|} \right|^\alpha \right\} \\
&= \exp\left\{ -\int_{\mathbb{S}^{|\mathbf{P}^{(1)} \times K|-1}} |\mathbf{t}^{(1)} \otimes s^{(1)}|^\alpha \Gamma^{(1)}(\mathrm{d}\, s^{(1)}) \right\}
\end{aligned}
$$

where $(x^{(1:K)}_\star)_{(c^{(0)},g^{(1)})} = \left[(x^{(k)}_{\star p^{(1)}})_{(c^{(0)},g^{(1)})} \right]_{\{(p^{(1)},k) \in [\mathbf{P}^{(1)} \times K]\}}$ and

$$
\begin{aligned}
\Gamma^{(1)} =& \|\sigma_b \mathbb{1}_{(\mathbf{P}^{(1)} \times K)}\|^\alpha \Psi^{(1)}(\mathbb{1}_{(\mathbf{P}^{(1)} \times K)}) \\
&+ \sum_{(c^{(0)},g^{(1)}) \in [C^{(0)} \times \mathbf{G}^{(1)}]} \|\sigma_\omega (x^{(1:K)}_\star)_{(c^{(0)},g^{(1)})}\|^\alpha \Psi^{(1)}\Big((x^{(1:K)}_\star)_{(c^{(0)},g^{(1)})} \Big)
\end{aligned}
$$

16

$\square$

SM A.2

We prove that for each $l = 2, \ldots, L$, $f_{(c^{(l)},:)}^{(l)(1:K)} | f_{(1:C,:)}^{(l-1)(1:K)} \sim \mathrm{St}_{\mathbf{P}^{(l)} \times K}(\alpha, \Gamma^{(l)})$, where

$$\Gamma_C^{(l)} = \|\sigma_b \mathbb{1}_{(\mathbf{P}^{(l)} \times K)}\|^\alpha \Psi^{(l)}\left(\mathbb{1}_{(\mathbf{P}^{(l)} \times K)}\right) +$$
$$+ \frac{1}{C} \sum_{(c^{(l-1)}, g^{(l)}) \in [C \times \mathbf{G}^{(l)}]} \|\sigma_\omega \phi(f_\star^{(l-1)(1:K)})_{(c^{(l-1)}, g^{(l)})}\|^\alpha \Psi^{(l)}\left(\phi(f_\star^{(l-1)(1:K)})_{(c^{(l-1)}, g^{(l)})}\right)$$

Proof. For $l \geq 2$, from definition (5) and assumption (3) we have that, for any $\mathbf{t}^{(l)} := [t_{p^{(l)}}^{(l)(k)}]_{\{(p^{(l)}, k) \in [\mathbf{P}^{(l)} \times K]\}} \in \mathbb{R}^{\mathbf{P}^{(l)} \times K}$, it holds

$$\varphi_{\left(f_{(c^{(l)},:)}^{(l)(1:K)} | f_{(1:C,:)}^{(l-1)(1:K)}\right)}(\mathbf{t}^{(l)})$$

$$= \mathbb{E}\left[\exp\left\{i\,\mathbf{t}^{(l)} \otimes f_{(c^{(l)},:)}^{(l)(1:K)} | f_{(1:C,:)}^{(l-1)(1:K)}\right\}\right]$$

$$= \mathbb{E}\left[\exp\left\{i\,\mathbf{t}^{(l)} \otimes \left(\frac{1}{C^{1/\alpha}} W_{(c^{(l)},:,:)}^{(l)} \overset{(\mathbf{P}^{(l)}, K)}{\otimes} \phi(f_\star^{(l-1)(1:K)}) + b_{c^{(l)}}^{(l)} \mathbb{1}_{(\mathbf{P}^{(l)} \times K)}\right)\right\}\right]$$

$$= \mathbb{E}\left[\exp\left\{i\,\frac{1}{C^{1/\alpha}} W_{(c^{(l)},:,:)}^{(l)} \overset{(\mathbf{P}^{(l)}, K)}{\otimes} \left(\mathbf{t}^{(l)} \overset{(C, \mathbf{G}^{(l)})}{\otimes} \phi(f_\star^{(l-1)(1:K)})\right) + \left(b_{c^{(l)}}^{(l)} \mathbf{t}^{(l)} \otimes \mathbb{1}_{(\mathbf{P}^{(l)} \times K)}\right)\right\}\right]$$

$$= \mathbb{E}\left[\exp\left\{b_{c^{(l)}}^{(l)} \mathbf{t}^{(l)} \otimes \mathbb{1}_{(\mathbf{P}^{(l)} \times K)}\right\}\right] \times$$

$$\times \prod_{(c^{(l-1)}, g^{(l)}) \in [C \times \mathbf{G}^{(l)}]} \mathbb{E}\left[\exp\left\{i\,\frac{1}{C^{1/\alpha}} W_{(c^{(l)}, c^{(l-1)}, g^{(l)})}^{(l)} \left(\mathbf{t}^{(l)} \otimes \phi(f_\star^{(l-1)(1:K)})_{(c^{(l-1)}, g^{(l)})}\right)\right\}\right]$$

$$= e^{-\sigma_b^\alpha |\mathbf{t}^{(l)} \otimes \mathbb{1}_{(\mathbf{P}^{(l)} \times K)}|^\alpha} \prod_{(c^{(l-1)}, g^{(l)}) \in [C \times \mathbf{G}^{(l)}]} e^{-\frac{\sigma_\omega^\alpha}{C}\left|\mathbf{t}^{(l)} \otimes \phi(f_\star^{(l-1)(1:K)})_{(c^{(l-1)}, g^{(l)})}\right|^\alpha}$$

$$= \exp\left\{-\sigma_b^\alpha |\mathbf{t}^{(l)} \otimes \mathbb{1}_{(\mathbf{P}^{(l)} \times K)}|^\alpha - \frac{\sigma_\omega^\alpha}{C} \sum_{(c^{(l-1)}, g^{(l)}) \in [C \times \mathbf{G}^{(l)}]} \left|\mathbf{t}^{(l)} \otimes \phi(f_\star^{(l-1)(1:K)})_{(c^{(l-1)}, g^{(l)})}\right|^\alpha\right\}$$

$$= \exp\left\{-\sigma_b^\alpha \|\mathbb{1}_{(\mathbf{P}^{(l)} \times K)}\|^\alpha \left|\mathbf{t}^{(l)} \otimes \frac{\mathbb{1}_{(\mathbf{P}^{(l)} \times K)}}{\|\mathbb{1}_{(\mathbf{P}^{(l)} \times K)}\|}\right|^\alpha +\right.$$

$$\left. -\frac{\sigma_\omega^\alpha}{C} \sum_{(c^{(l-1)}, g^{(l)}) \in [C \times \mathbf{G}^{(l)}]} \|\phi(f_\star^{(l-1)(1:K)})_{(c^{(l-1)}, g^{(l)})}\|^\alpha \left|\mathbf{t}^{(l)} \otimes \frac{(f_\star^{(l-1)(1:K)})_{(c^{(l-1)}, g^{(l)})}}{\|\phi(f_\star^{(l-1)(1:K)})_{(c^{(l-1)}, g^{(l)})}\|}\right|^\alpha\right\}$$

$$= \exp\left\{-\int_{\mathbb{S}^{|\mathbf{P}^{(l)} \times K|-1}} |\mathbf{t}^{(l)} \otimes s^{(l)}|^\alpha \Gamma_C^{(l)}(d\,s^{(l)})\right\}$$

where

$$\Gamma_C^{(l)} = \|\sigma_b \mathbb{1}_{(\mathbf{P}^{(l)} \times K)}\|^\alpha \Psi^{(l)}\left(\mathbb{1}_{(\mathbf{P}^{(l)} \times K)}\right) +$$
$$+ \frac{1}{C} \sum_{(c^{(l-1)}, g^{(l)}) \in [C \times \mathbf{G}^{(l)}]} \|\sigma_\omega \phi(f_\star^{(l-1)(1:K)})_{(c^{(l-1)}, g^{(l)})}\|^\alpha \Psi^{(l)}\left(\phi(f_\star^{(l-1)(1:K)})_{(c^{(l-1)}, g^{(l)})}\right)$$

with $(f_\star^{(l-1)(1:K)})_{(c^{(l-1)}, g^{(l)})} = \left[(f_{\star p^{(l)}}^{(l-1)(k)})_{(c^{(l-1)}, g^{(l)})}\right]_{\{(p^{(l)}, k) \in [\mathbf{P}^{(l)} \times K]\}}$. $\square$

SM B

We prove that for each $l \in [L]$, $f^{(l)(1:K)} = f^{(l)}(x^{(1:K)}, C) \xrightarrow{d} \bigotimes_{c^{(l)}=1}^{\infty} \mathrm{St}_{\mathbf{P}^{(l)} \times K}(\alpha, \Gamma_{\infty}^{(l)})$ as $C \to \infty$.

The symbol $\bigotimes$ here denotes the product measure. The proof follows by the Cramér-Wold theorem for finite-dimensional projection of $f^{(l)(1:K)} = f^{(l)}(x^{(1:K)}, C)$ for which it is sufficient to prove the large C asymptotic behavior of any linear combination of the $f_{(c^{(l)},:)}^{(l)(1:K)}$'s for $c^{(l)} \in \mathcal{L} \subset \mathbb{N}$. See, e.g. Billingsley [1999] for details.

Proof. Following the notation of Matthews et al. [2018], consider a finite linear combination of the function values without the bias, i.e. fix $z = (z_{c^{(l)}})_{\{c^{(l)} \in \mathcal{L}\}}$ and define

$$\mathcal{T}^{(l)}(\mathcal{L}, z, x^{(1:K)}, C^{(l-1)}) = \sum_{c^{(l)} \in \mathcal{L}} z_{c^{(l)}} [f_{(c^{(l)},:)}^{(l)(1:K)} - b_{c^{(l)}}^{(l)} \mathbb{1}_{(\mathbf{P}^{(l)} \times K)}].$$

The case $l = 1$ is easy since it does not depend on C, indeed we get

$$\mathcal{T}^{(1)}(\mathcal{L}, z, x^{(1:K)}, C^{(0)}) = \sum_{c^{(1)} \in \mathcal{L}} z_{c^{(1)}} W_{(c^{(1)},:,:)}^{(1)} \overset{(\mathbf{P}^{(1)}, K)}{\otimes} x_{\star}^{(1:K)}$$

and, following the same steps as in Theorem 3, for any $\mathbf{t}^{(1)} := [t_{p^{(1)}}^{(1)(k)}]_{\{(p^{(1)},k) \in [\mathbf{P}^{(1)} \times K]\}} \in \mathbb{R}^{\mathbf{P}^{(1)} \times K}$, called $\|z\|^{\alpha} = \sum_{c^{(1)} \in \mathcal{L}} |z_{c^{(1)}}|^{\alpha}$, we get

$$
\begin{aligned}
&\varphi_{\left(T^{(1)}(\mathcal{L}, z, x^{(1:K)}, C^{(0)})\right)}(\mathbf{t}^{(1)}) \\
&= \mathbb{E}\left[\exp\left\{i\,\mathbf{t}^{(1)} \otimes T^{(1)}(\mathcal{L}, z, x^{(1:K)}, C^{(0)})\right\}\right] \\
&= \mathbb{E}\left[\exp\left\{i\,\mathbf{t}^{(1)} \otimes \left(\sum_{c^{(1)} \in \mathcal{L}} z_{c^{(1)}} W_{(c^{(1)},:,:)}^{(1)} \overset{(\mathbf{P}^{(1)}, K)}{\otimes} x_{\star}^{(1:K)}\right)\right\}\right] \\
&= \mathbb{E}\left[\exp\left\{i\sum_{c^{(1)} \in \mathcal{L}} z_{c^{(1)}} W_{(c^{(1)},:,:)}^{(1)} \overset{(\mathbf{P}^{(1)}, K)}{\otimes} \left(\mathbf{t}^{(1)} \overset{(C^{(0)}, \mathbf{G}^{(1)})}{\otimes} x_{\star}^{(1:K)}\right)\right\}\right] \\
&= \prod_{(c^{(1)}, c^{(0)}, g^{(1)}) \in \mathcal{L} \times [C^{(0)} \times \mathbf{G}^{(1)}]} \mathbb{E}\left[\exp\left\{i\,z_{c^{(1)}} W_{(c^{(1)}, c^{(0)}, g^{(1)})}^{(1)} \mathbf{t}^{(1)} \otimes (x_{\star}^{(1:K)})_{(c^{(0)}, g^{(1)})}\right\}\right] \\
&= \prod_{(c^{(1)}, c^{(0)}, g^{(1)}) \in \mathcal{L} \times [C^{(0)} \times \mathbf{G}^{(1)}]} \exp\left\{-(|z_{c^{(1)}}|\sigma_{\omega})^{\alpha}\left|\mathbf{t}^{(1)} \otimes (x_{\star}^{(1:K)})_{(c^{(0)}, g^{(1)})}\right|^{\alpha}\right\} \\
&= \exp\left\{-\sigma_{\omega}^{\alpha}\|z\|^{\alpha} \sum_{(c^{(0)}, g^{(1)}) \in [C^{(0)} \times \mathbf{G}^{(1)}]} \left|\mathbf{t}^{(1)} \otimes (x_{\star}^{(1:K)})_{(c^{(0)}, g^{(1)})}\right|^{\alpha}\right\} \\
&= \exp\left\{-\sigma_{\omega}^{\alpha}\|z\|^{\alpha} \sum_{(c^{(0)}, g^{(1)}) \in [C^{(0)} \times \mathbf{G}^{(1)}]} \|(x_{\star}^{(1:K)})_{(c^{(0)}, g^{(1)})}\|^{\alpha}\left|\mathbf{t}^{(1)} \otimes \frac{(x_{\star}^{(1:K)})_{(c^{(0)}, g^{(1)})}}{\|(x_{\star}^{(1:K)})_{(c^{(0)}, g^{(1)})}\|}\right|^{\alpha}\right\} \\
&= \exp\left\{-\int_{\mathbb{S}^{|\mathbf{P}^{(1)} \times K| - 1}} |\mathbf{t}^{(1)} \otimes s^{(1)}|^{\alpha} \Delta_C^{(1)}(\mathrm{d}\,s^{(1)})\right\}
\end{aligned}
$$

where $\Delta_C^{(1)}$ coincides with $\Gamma^{(1)}$ just replacing $\sigma_b \leftarrow 0$ and $\sigma_{\omega} \leftarrow \sigma_{\omega}\|z\|$. Thus since the characteristic function does not depend on C we get (as $C \to \infty$)

$$\mathcal{T}^{(1)}(\mathcal{L}, z, x^{(1:K)}, C^{(0)}) \xrightarrow{d} \mathrm{St}_{\mathbf{P}^{(1)} \times K}(\alpha, \Delta_{\infty}^{(1)})$$

where

$$\Delta^{(1)} = \Delta_{\infty}^{(1)} = \|z\|^{\alpha} \sum_{(c^{(0)}, g^{(1)}) \in [C^{(0)} \times \mathbf{G}^{(1)}]} \|\sigma_{\omega}(x_{\star}^{(1:K)})_{(c^{(0)}, g^{(1)})}\|^{\alpha} \Psi^{(1)}\left((x_{\star}^{(1:K)})_{(c^{(0)}, g^{(1)})}\right)$$

18

For $l = 2, \dots, L$,

$$\mathcal{T}^{(l)}(\mathcal{L}, z, x^{(1:K)}, C) = \sum_{c^{(l)} \in \mathcal{L}} \frac{z_{c^{(l)}}}{C^{1/\alpha}} W^{(l)}_{(c^{(l)}, :, :)} \overset{(\mathbf{P}^{(l)}, K)}{\otimes} \phi(f_{\star}^{(l-1)(1:K)})$$

and, following the same steps as in Theorem 4, for any $\mathbf{t}^{(l)} := [t^{(l)(k)}_{p^{(l)}}]_{\{(p^{(l)}, k) \in [\mathbf{P}^{(l)} \times K]\}} \in \mathbb{R}^{\mathbf{P}^{(l)} \times K}$, called $\|z\|^{\alpha} = \sum_{c^{(l)} \in \mathcal{L}} |z_{c^{(l)}}|^{\alpha}$, we get

$$\varphi_{\left(T^{(l)}(\mathcal{L}, z, x^{(1:K)}, C) | f^{(l-1)(1:K)}_{(1:C, :)}\right)}(\mathbf{t}^{(l)})$$

$$= \mathbb{E}\left[\exp\left\{ i\, \mathbf{t}^{(l)} \otimes T^{(l)}(\mathcal{L}, z, x^{(1:K)}, C) | f^{(l-1)(1:K)}_{(1:C, :)}\right\}\right]$$

$$= \mathbb{E}\left[\exp\left\{ i\, \mathbf{t}^{(l)} \otimes \Big(\sum_{c^{(l)} \in \mathcal{L}} \frac{z_{c^{(l)}}}{C^{1/\alpha}} W^{(l)}_{(c^{(l)}, :, :)} \overset{(\mathbf{P}^{(l)}, K)}{\otimes} \phi(f_{\star}^{(l-1)(1:K)}) \Big)\right\}\right]$$

$$= \mathbb{E}\left[\exp\left\{ i \sum_{c^{(l)} \in \mathcal{L}} \frac{z_{c^{(l)}}}{C^{1/\alpha}} W^{(l)}_{(c^{(l)}, :, :)} \overset{(\mathbf{P}^{(l)}, K)}{\otimes} \left(\mathbf{t}^{(l)} \overset{(C, \mathbf{G}^{(l)})}{\otimes} \phi(f_{\star}^{(l-1)(1:K)}) \right)\right\}\right]$$

$$= \prod_{(c^{(l)}, c^{(l-1)}, g^{(l)}) \in \mathcal{L} \times [C \times \mathbf{G}^{(l)}]} \mathbb{E}\left[\exp\left\{ i \frac{z_{c^{(l)}}}{C^{1/\alpha}} W^{(l)}_{(c^{(l)}, c^{(l-1)}, g^{(l)})} \mathbf{t}^{(l)} \otimes \phi(f_{\star}^{(l-1)[1:K]})_{(c^{(l-1)}, g^{(l)})} \right\}\right]$$

$$= \prod_{(c^{(l)}, c^{(l-1)}, g^{(l)}) \in \mathcal{L} \times [C \times \mathbf{G}^{(l)}]} \exp\left\{ - \frac{(|z_{c^{(l)}}| \sigma_{\omega})^{\alpha}}{C} \left| \mathbf{t}^{(l)} \otimes \phi(f_{\star}^{(l-1)(1:K)})_{(c^{(l-1)}, g^{(l)})} \right|^{\alpha} \right\}$$

$$= \exp\left\{ - \frac{\sigma_{\omega}^{\alpha}}{C} \|z\|^{\alpha} \sum_{(c^{(l-1)}, g^{(l)}) \in [C \times \mathbf{G}^{(l)}]} \left| \mathbf{t}^{(l)} \otimes \phi(f_{\star}^{(l-1)(1:K)})_{(c^{(l-1)}, g^{(l)})} \right|^{\alpha} \right\}$$

$$= \exp\left\{ - \frac{\sigma_{\omega}^{\alpha}}{C} \|z\|^{\alpha} \sum_{(c^{(l-1)}, g^{(l)}) \in [C \times \mathbf{G}^{(l)}]} \|\phi(f_{\star}^{(l-1)(1:K)})_{(c^{(l-1)}, g^{(l)})}\|^{\alpha} \times \right.$$

$$\left. \times \left| \mathbf{t}^{(l)} \otimes \frac{(f_{\star}^{(l-1)(1:K)})_{(c^{(l-1)}, g^{(l)})}}{\|\phi(f_{\star}^{(l-1)(1:K)})_{(c^{(l-1)}, g^{(l)})}\|} \right|^{\alpha} \right\}$$

$$= \exp\left\{ - \int_{\mathbb{S}^{|\mathbf{P}^{(l)} \times K| - 1}} |\mathbf{t}^{(l)} \otimes s^{(l)}|^{\alpha} \Delta_C^{(l)}(\mathrm{d}\, s^{(l)}) \right\}$$

where $\Delta_C^{(l)}$ coincides with $\Gamma_C^{(l)}$ just replacing $\sigma_b \leftarrow 0$ and $\sigma_{\omega} \leftarrow \sigma_{\omega} \|z\|$. Now, proceeding as in Theorem 1, we get the weak limit as $C \to +\infty$, i.e.

$$\mathcal{T}^{(l)}(\mathcal{L}, z, x^{(1:K)}, C) \overset{d}{\to} \mathrm{St}_{\mathbf{P}^{(l)} \times K}(\alpha, \Delta_{\infty}^{(l)})$$

where

$$\Delta_{\infty}^{(l)} = \|z\|^{\alpha} \int \sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} \|\sigma_{\omega} \phi(f_{g^{(l)}})\|^{\alpha} \Psi^{(l)}\Big(\phi(f_{g^{(l)}}) \Big) q^{(l-1)}(\mathrm{d}\, f_{\{g^{(l)} \in [\mathbf{G}^{(l)}]\}})$$

This completes the proof. $\qquad \square$

SM C

Focus the attention on the last layer L. We found that $f^{(L)(1:K)} \overset{d}{\to} f_{\infty}^{(L)(1:K)}$, i.e. a convergence of a sequence of $\mathbb{R}^{\infty \times \mathbf{P}^{(L)} \times K}$-valued random variables. To gather information on the positions $\mathbf{P}^{(L)}$ we consider a linear combination with respect to $\mathbf{P}^{(L)}$, i.e. we project the $\infty \times \mathbf{P}^{(L)} \times K$ dimensional vector $f^{(L)(1:K)} = f^{(L)}(x^{(1:K)}, C)$ into a $\infty \times K$ dimensional vectors, and we take the limit as $C \to \infty$. More precisely, for $l \in [L]$, fix $\mathbf{u} \in \mathbb{R}^{\mathbf{P}^{(l)}}$ such that $\mathbf{u} \otimes \mathbb{1}_{(\mathbf{P}^{(l)})} = 1$ and define the

19

transformation $T_{\mathbf{u}}^{(l)} : \mathbb{R}^{\infty \times \mathbf{P}^{(l)} \times K} \to \mathbb{R}^{\infty \times K}$, $(a, b, c) \mapsto (a, \mathbf{u} \otimes b, c)$ (in other words $T_{\mathbf{u}}^{(l)} \equiv \mathbf{u} \underset{\mathbf{P}^{(l)}}{\otimes}$).

We want to establish the convergence of $T_{\mathbf{u}}^{(l)}(f^{(l)(1:K)})$ as $C \to \infty$.

For $l = 1$ we get

$$
\begin{aligned}
T_{\mathbf{u}}^{(1)}(f^{(1)(1:K)}) &= T_{\mathbf{u}}^{(1)}\left(W^{(1)} \overset{(\mathbf{P}^{(1)}, K)}{\boxdot} x_{\star}^{(1:K)} + b^{(1)} \triangle \mathbb{1}_{(\mathbf{P}^{(1)} \times K)}\right) \\
&= W^{(1)} \overset{(K)}{\boxdot} (\mathbf{u} \underset{\mathbf{P}^{(1)}}{\otimes} x_{\star}^{(1:K)}) + b^{(1)} \triangle \mathbb{1}_{(K)}
\end{aligned}
$$

and for $l > 1$,

$$
\begin{aligned}
T_{\mathbf{u}}^{(l)}(f^{(l)(1:K)}) &= T_{\mathbf{u}}^{(l)}\left(\frac{1}{C^{1/\alpha}} W^{(l)} \overset{(\mathbf{P}^{(l)}, K)}{\boxdot} \phi(f_{\star}^{(l-1)(1:K)}) + b^{(l)} \triangle \mathbb{1}_{(\mathbf{P}^{(l)} \times K)}\right) \\
&= \frac{1}{C^{1/\alpha}} W^{(l)} \overset{(K)}{\boxdot} (\mathbf{u} \underset{\mathbf{P}^{(l)}}{\otimes} \phi(f_{\star}^{(l-1)(1:K)})) + b^{(l)} \triangle \mathbb{1}_{(K)}
\end{aligned}
$$

Following the same steps of Theorem 2 we get that

$$
T_{\mathbf{u}}^{(l)}(f^{(l)(1:K)}) \overset{d}{\to} \bigotimes_{c^{(l)}=1}^{\infty} \mathrm{St}_K(\alpha, \Gamma_{\infty}^{(l)}(\mathbf{u}))
$$

where

$$
\Gamma_{\infty}^{(l)}(\mathbf{u}) = \|\sigma_b \mathbb{1}_{(K)}\|^{\alpha} \Psi^{(l)}\left(\mathbb{1}_{(K)}\right) + \int \sum_{g^{(l)} \in [\mathbf{G}^{(l)}]} \|\sigma_{\omega} \mathbf{u} \underset{\mathbf{P}^{(l)}}{\otimes} \phi(f_{g^{(l)}})\|^{\alpha} \times
$$

$$
\times \mathcal{D}^{(l)}\left(\mathbf{u} \underset{\mathbf{P}^{(l)}}{\otimes} \phi(f_{g^{(l)}})\right) q^{(l-1)}(\mathrm{d}\, f_{\{g^{(l)} \in [\mathbf{G}^{(l)}]\}})
$$

where $\mathcal{D}^{(l)} : \mathbb{R}^K \to \mathbb{R}$

$$
\mathcal{D}^{(l)}(z) := \begin{cases} \frac{1}{2}\delta\left(\frac{z}{\|z\|}\right) + \frac{1}{2}\delta\left(-\frac{z}{\|z\|}\right) & 0 \neq z \in \mathbb{R}^K \\ 0 & 0 = z \in \mathbb{R}^K \end{cases}
$$

$f_{g^{(l)}} \in \mathbb{R}^{\mathbf{P}^{(l)} \times K}$ and $q^{(l)} = \mathrm{St}_{\mathbf{P}^{(l)} \times K}(\alpha, \Gamma_{\infty}^{(l)})$ for $l \in [L]$, being $\Gamma_{\infty}^{(l)}$ defined in theorem 1.