Double Continuum Limit of Deep Neural Networks: Supplementary Material

Sho Sonoda Noboru Murata

A Ridgelet Transform of Truncated Autoencoder

Let $\varepsilon \to 0$, then the ridgelet transform of the truncated autoencoder $\mathsf{id}_{r,\varepsilon}$ is given by

$$\mathcal{R}_\rho[\mathsf{id}_{r,0}](a,b) = -\frac{A_{m-1}}{2(m+1)} \int_{|p|<r} (r^2 - p^2)^{\frac{m-1}{2}} \left\{ \frac{2}{m-1} p^2 + r^2 \right\} \overline{\rho'(|a|p - b)} a \mathrm{d}p \tag{A.1}$$

$$\approx -KC(a,b) a \overline{\rho'(-b)}, \tag{A.2}$$

where constants A_{m-1} and K are defined in below.

Proof. Let $\varepsilon \to 0$, then the connecting annulus $\mathbb{B}(0; r+\varepsilon) \setminus \mathbb{B}(0; r)$ vanishes as below:

$$\begin{aligned}
\mathcal{R}_\rho[\mathsf{id}_{r,\varepsilon}](a,b) &= -a \mathcal{R}_{\rho'}[V_{r,\varepsilon}](a,b) \\
&\to -aC(a,b) \int_{\mathbb{B}^m(r)} \frac{1}{2} |x|^2 \overline{\rho'(a \cdot x - b)} \mathrm{d}x \\
&= -a \mathcal{R}_{\rho'}[V_{r,0}](a,b).
\end{aligned}$$

Hence, we omit considering the annulus.

In the followings, we use a spherical coordinate defined by

$$u := a/|a|, \quad \alpha := 1/|a|, \quad \beta := b/|a|,$$

where $u \in \mathbb{S}^{m-1}$ denotes the direction, $\alpha \in \mathbb{R}_+$ the scale, and $\beta \in \mathbb{R}$ the (scaled) shift parameters respectively.

The ridgelet transform in the spherical coordinate (Sonoda & Murata, 2017) is given by

$$\mathcal{R}_\rho f(u/\alpha, \beta/\alpha) = C(u/\alpha, \beta/\alpha) \int_{\mathbb{R}} \mathrm{R}f(u,p) \overline{\rho_\alpha(p - \beta)} \mathrm{d}p,$$

where $\mathrm{R}f(u,p)$ denotes the Radon transform

$$\mathrm{R}f(u,p) := \int_{(\mathbb{R}u)^\perp} f(pu + y) \mathrm{d}y,$$

of function $f \in L^1(\mathbb{R}^m)$ at direction $u \in \mathbb{S}^{m-1}$ and position $p \in \mathbb{R}$; and

$$\rho_\alpha(p) := \rho(p/\alpha).$$

The Radon transform $\mathrm{R}[V_{r,0}](u,p)$ for $|p| < r$ is calculated as follows. Because $V_{r,\varepsilon}$ is a radial function, $\mathrm{R}[V_{r,0}](u,p)$ does not depends on direction u. Hence, it is sufficient to consider a spacial case when $(\mathbb{R}u)^\perp = \mathbb{R}^{m-1}$. Therefore,

$$\begin{aligned}
\mathrm{R}[V_{r,0}](u,p) &= \int_{\mathbb{R}^{m-1}} V_{r,0}(pu + y) \mathrm{d}y, \quad u \perp y \\
&= \int_{\mathbb{R}^{m-1}} \frac{1}{2} |pu + y|^2 \mathbf{1}_{\mathbb{B}^m(0;r)}(pu + y) \mathrm{d}y \\
&= \frac{1}{2} \int_{\mathbb{B}^{m-1}\left(0; \sqrt{r^2 - p^2}\right)} \left\{ p^2 + |y|^2 \right\} \mathrm{d}y
\end{aligned} \tag{A.3}$$

where the third equation follows by the orthogonality $|pu+y|_m^2 = p^2 + |y|_{m-1}^2$, and a geometric consideration as below:

$$\int_{\mathbb{R}^{m-1}} [\,\cdot\,] \mathbf{1}_{\mathbb{B}^m(0;r)}(pu+y)\mathrm{d}y = \int_{\mathbb{R}^{m-1}} [\,\cdot\,] \mathbf{1}_{\mathbb{B}^m(-pu;r)}(y)\mathrm{d}y$$
$$= \int_{\mathbb{R}^{m-1}\cap\mathbb{B}^m(-pu;r)} [\,\cdot\,]\mathrm{d}y$$
$$= \int_{\mathbb{B}^{m-1}(0;\sqrt{r^2-p^2})} [\,\cdot\,]\mathrm{d}y.$$

The first integral in (A.3) is calculated as below:

$$\int_{\mathbb{B}^{m-1}\left(0;\sqrt{r^2-p^2}\right)} p^2 \mathrm{d}y = p^2 \operatorname{vol}\left[\mathbb{B}^{m-1}(0;\sqrt{r^2-p^2})\right]$$
$$= \frac{\pi^{\frac{m-1}{2}}}{2\Gamma\left(\frac{m-1}{2}+1\right)} p^2 (r^2-p^2)^{\frac{m-1}{2}}. \tag{A.4}$$

The second integral in (A.3) is calculated as below:

$$\int_{\mathbb{B}^{m-1}\left(0;\sqrt{r^2-p^2}\right)} |y|^2 \mathrm{d}y = \int_{\mathbb{S}^{m-2}} \int_0^{\sqrt{r^2-p^2}} |\rho\omega|^2 \rho^{m-2} \mathrm{d}\rho\mathrm{d}\omega$$
$$= \int_{\mathbb{S}^{m-2}} \mathrm{d}\omega \int_0^{\sqrt{r^2-p^2}} \rho^m \mathrm{d}\rho$$
$$= \frac{\pi^{\frac{m-1}{2}}}{(m+1)\Gamma\left(\frac{m-1}{2}\right)} (r^2-p^2)^{\frac{m+1}{2}}. \tag{A.5}$$

Hence, by combining the first and second integrals, we have

$$\mathrm{R}[V_{r,0}](u,p) = \begin{cases} \frac{A_{m-1}}{2(m+1)} (r^2-p^2)^{\frac{m-1}{2}} \left\{\frac{2}{m-1}p^2 + r^2\right\} & |p| < r \\ 0 & |p| \geq r, \end{cases} \tag{A.6}$$

where $A_{m-1} := \frac{2\pi^{\frac{m-1}{2}}}{\Gamma\left(\frac{m-1}{2}\right)}$ is the surface area of $\mathbb{S}^{m-1}$.

The ridgelet transform $\mathcal{R}_{\rho'}[V_{r,0}]$ is given by

$$\mathcal{R}_{\rho'}[V_{r,0}](u/\alpha, \beta/\alpha) = \int_{|p|<r} k(p)\overline{\rho'_\alpha(p-\beta)}\mathrm{d}p, \tag{A.7}$$

where we define

$$k(p) := \mathrm{R}[V_{r,0}](u,p).$$

Recall that $\mathrm{R}[V_{r,0}](u,p)$ does not depend on direction u, thus the definiton of k is reasonable. According to (A.6), k is a compactly supported bump function. As a consequence, k is summable and thus the integral

$$K := \int_{\mathbb{R}} k(p)\mathrm{d}p$$

always exists. Recall that the convolution results in smoothing. That is,

$$\int_{|p|<r} k(p)\overline{\rho'_\alpha(p-\beta)}\mathrm{d}p \approx K\overline{\rho'_\alpha(-\beta)}. \tag{A.8}$$

To sum up, we have presented the followings.

$$\mathcal{R}_\rho[\mathsf{id}_{r,0}](a,b) = -a\mathcal{R}_{\rho'}[V_{r,0}](a,b) \approx -KC(a,b)a\overline{\rho'(-b)}. \qquad \square$$

References

Sonoda, Sho and Murata, Noboru. Neural network with unbounded activation functions is universal approximator. *Applied and Computational Harmonic Analysis*, 43(2):233–268, 2017.