A PROOFS

Theorem 2 (Kernel invariance under orthogonal representations)**.** *Let G be a group and ρ_X an orthogonal representation of G acting on the input space X. Under the representation ρ_X, the neural tangent kernel Θ as defined in (3) as well as the NNGP kernel $\mathcal{K}$ as defined in (10) of a neural network satisfying the assumptions above are invariant:*

$$\Theta(x, x') = \Theta(\rho_X(g)x, \rho_X(g)x')\,, \tag{16}$$

$$\mathcal{K}(x, x') = \mathcal{K}(\rho_X(g)x, \rho_X(g)x')\,. \tag{17}$$

for all $g \in G$ and $x, x' \in X$.

Proof. To simplify notation, we will use the shorthand notation

$$O = \rho_X(g) \tag{31}$$

for the orthogonal representation matrix corresponding to the action of group element g on the input space X. This orthogonal matrix can be absorbed by redefining the parameters which multiply the input. For notational simplicity, we assume an MLP without biases in the following, but the proof immediately generalizes to the case of dependency on linear transformations of the input as stated in the assumptions.

Redefining the parameters $w^{(1)}$ of the first layer of the neural network $f_w : X \to Y$ yields

$$f_w(Ox) = f_{w'}(x)\,, \tag{32}$$

where we have defined the new weights of layer l as

$$w'^{(l)} = \begin{cases} w^{(1)}O, & \text{if } l = 1 \\ w^{(l)}, & \text{otherwise}\,. \end{cases} \tag{33}$$

We can use this result to rewrite the gradient of the network with respect to the parameters of the first layer

$$\frac{\partial}{\partial w^{(1)}_{ij}} f_w(Ox) = \frac{\partial}{\partial w^{(1)}_{ij}} f_{w'}(x) = \frac{\partial w'^{(1)}_{mn}}{\partial w^{(1)}_{ij}} \frac{\partial}{\partial w'^{(1)}_{mn}} f_{w'}(x) = O_{jn} \frac{\partial}{\partial w'^{(1)}_{in}} f_{w'}(x)\,, \tag{34}$$

where here and in the following we use the Einstein summation convention. It is convenient to define

$$O^{(l)} = \begin{cases} O, & \text{if } l = 1 \\ \mathbb{I}, & \text{otherwise}\,, \end{cases} \tag{35}$$

such that the above gradient relation can be generalized to

$$\frac{\partial f_w(Ox)}{\partial w^{(l)}_{ij}} = O^{(l)}_{jn} \frac{\partial f_{w'}(x)}{\partial w'^{(l)}_{in}}\,. \tag{36}$$

Since the representation is orthogonal, it holds that for all $l \in \{1, \dots, L\}$

$$(O^{(l)})^\top O^{(l)} = \mathbb{I} \iff (O^{(l)})_{jn} O^{(l)}_{jm} = \delta_{nm}\,, \tag{37}$$

where δ_{nm} is the Kronecker symbol.

The neural tangent kernel (3) involves a sum over all layers and can thus be rewritten as

$$\Theta(Ox, Oy) = \sum_{l=1}^{L} \mathbb{E}_{w\sim p}\left[\frac{\partial f_w(Ox)}{\partial w^{(l)}_{ij}} \frac{\partial f_w(Oy)}{\partial w^{(l)}_{ij}}\right] \tag{38}$$

$$= \sum_{l=1}^{L} \mathbb{E}_{w\sim p}\left[\frac{\partial f_w(Ox)}{\partial w^{(l)}_{ij}} \frac{\partial f_w(Oy)}{\partial w^{(l)}_{ij}}\right] \tag{39}$$

$$= \sum_{l=1}^{L} \mathbb{E}_{w\sim p}\left[\frac{\partial f_{w'}(x)}{\partial w'^{(l)}_{in}} \frac{\partial f_{w'}(y)}{\partial w'^{(l)}_{im}} (O^{(l)})_{jn} O^{(l)}_{jm}\right] \tag{40}$$

$$= \sum_{l=1}^{L} \mathbb{E}_{w\sim p}\left[\frac{\partial f_{w'}(x)}{\partial w'^{(l)}_{in}} \frac{\partial f_{w'}(y)}{\partial w'^{(l)}_{in}}\right]\,, \tag{41}$$