Supplementary Material for NIPS 2009 Paper #978, "On Invariance in Hierarchical Models"

Jake Bouvrie, Lorenzo Rosasco, and Tomaso Poggio
Center for Biological and Computational Learning
Massachusetts Institute of Technology
Cambridge, MA USA
{jvb,lrosasco}@mit.edu, tp@ai.mit.edu

This supplementary appendix to the paper entitled "On Invariance in Hierarchical Models" (NIPS 2009) [1] contains all of the proofs omitted from the body of the paper. It is available online for downloading at `http://cbcl.mit.edu/publications/ps/978_supplement.pdf`.

The following definition is reproduced here for convenience:

Definition 1 (Compatible Sets). *The subsets $\tilde{R} \subset R$ and $\tilde{T} \subset T$ are* compatible *if all of the following conditions hold:*

1. *For each $r \in \tilde{R}$, $r_v\tilde{T} = \tilde{T}r_u$. When $r_u = r_v$ for all $r \in R$, this means that normalizer of $\tilde{T}$ in $\tilde{R}$ is $\tilde{R}$.*
2. *Left transformations r_v never take a point in v outside of v, and right transformations r_u never take a point in u/v outside of u/v (respectively):*
$$imA_{r_v} \circ \iota_v \subseteq v, \qquad imA_{r_u} \circ \iota_u \subseteq u, \qquad imA_{r_u} \circ \iota_v \subseteq v,$$
for all $r \in \tilde{R}$.
3. *Translations never take a point in u outside of v:*
$$imA_t \circ \iota_u \subseteq v$$
for all $t \in \tilde{T}$.

Theorem 1. *Given any function $\Psi : \mathcal{B}(\mathbb{R}_{++}) \to \mathbb{R}_{++}$, if the initial kernel satisfies $\widehat{K}_1(f, f \circ r) = 1$ for all $r \in \mathcal{R}$, $f \in \mathrm{Im}(v_1)$, then*
$$\widehat{N}_m(f) = \widehat{N}_m(f \circ r),$$
for all $r \in \mathcal{R}$, $f \in \mathrm{Im}(v_m)$ and $m \leq n$.

Proof. The proof is by induction. The base case is true by assumption. The inductive hypothesis is that $\widehat{K}_{m-1}(u, u \circ r) = 1$ for any $u \in \mathrm{Im}(v_{m-1})$. This means that $F(H \circ r) = F(H)$. Assumption 1 (in the full paper) states that $r \circ H = \pi(H) \circ r = H \circ r$, with π onto. Combining the inductive hypothesis and the Assumption, we see that for all $q \in Q_{m-1}$, $N_m(f \circ r)(q) = (\Psi \circ F)(r \circ H) = (\Psi \circ F)(H \circ r) = (\Psi \circ F)(H) = N_m(f)(q)$. $\square$

Proposition 1. *Let $\Gamma \subseteq T$ be a given set of translations, and assume the following: (1) $G \cong T \rtimes R$, (2) For each $r \in R$, $r = r_u = r_v$, (3) $\tilde{R}$ is a subgroup of R. Then Condition (1) of Definition 1 is satisfied if and only if $\tilde{T}$ can be expressed as a union of orbits of the form*
$$\tilde{T} = \bigcup_{t \in \Gamma} C_{\tilde{R}}(t)\,. \tag{1}$$

Proof. We first show that for $\tilde{T}$ of the form above, Condition (1) of Definition 1 is satisfied. Combining the first two assumptions, we have for all $t \in T$, $r \in \tilde{R}$ that $r_v t r_u^{-1} = rtr^{-1} \in T$. Then for each $r \in \tilde{R}$,

$$\begin{aligned} r_v \tilde{T} r_u^{-1} = r\tilde{T}r^{-1} &= r\left(\bigcup_{t\in\Gamma} C_{\tilde{R}}(t)\right) r^{-1} = r\left(\bigcup_{t\in\Gamma} \{\tilde{r}t\tilde{r}^{-1} \in T \mid \tilde{r} \in \tilde{R}\}\right) r^{-1} \\ &= \bigcup_{t\in\Gamma} \{r\tilde{r}t\tilde{r}^{-1}r^{-1} \in T \mid \tilde{r} \in \tilde{R}\} = \bigcup_{t\in\Gamma} \{r'tr'^{-1} \in T \mid r' \in \tilde{R}\} = \tilde{T} \end{aligned}$$

where the last equality follows since $r' \equiv r\tilde{r} \in \tilde{R}$, and $gG = G$ for any group G and $g \in G$ because $gG \subseteq G$ combined with the fact that $Gg^{-1} \subseteq G \Rightarrow Gg^{-1}g \subseteq Gg \Rightarrow G \subseteq Gg$ giving $gG = G$. So the condition is verified. Suppose now that Condition (1) is satisfied, but $\tilde{T}$ is not a union of orbits for the action of conjugation. Then there is a $t' \in \tilde{T}$ such that t' *cannot* be expressed as $t' = rtr^{-1}$ for some $t \in \tilde{T}$. Hence $r\tilde{T}r^{-1} \subset \tilde{T}$. But this contradicts the assumption that Condition (1) is satisfied, so $\tilde{T}$ must be of the form shown in Equation (1). □

Lemma 1. *For each $m \in M$, $t_a \in T$,*

$$mt_a = t_b m$$

for some unique element $t_b \in T$.

Proof. The group of isometries of the plane is generated by translations and orthogonal operators, so we can write an element $m \in M$ as $m = t_v\varphi$. Now define the homomorphism $\pi : M \to O_2$, sending $t_v\varphi \mapsto \varphi$. Clearly T is the kernel of π and is therefore a normal subgroup of M. Furthermore, $M = T \rtimes O_2$. So we have that, for all $m \in M$, $mt_am^{-1} = t_b$, for some element $t_b \in T$. Denote by $\varphi(v)$ the operation of $\varphi \in O_2$ on a vector $v \in \mathbb{R}^2$ given by the standard matrix representation $R : O_2 \to \mathsf{GL}_2(\mathbb{R})$ of O_2. Then $\varphi t_a = t_b\varphi$ with $b = \varphi(a)$ since $\varphi \circ t_a(x) = \varphi(x+a) = \varphi(x) + \varphi(a) = t_b\varphi(x)$. So for arbitrary isometries m, we have that

$$mt_am^{-1} = (t_v\varphi)t_a(\varphi^{-1}t_{-v}) = t_vt_b\varphi\varphi^{-1}t_{-v} = t_b,$$

where $b = \varphi(a)$. Since the operation of φ is bijective, b is unique, and $mTm^{-1} = T$ for all $m \in M$. □

Proposition 2. *Let H be the set of translations associated to an arbitrary layer of the hierarchical feature map and define the injective map $\tau : H \to T$ by $h_a \mapsto t_a$, where a is a parameter characterizing the translation. Set $\Gamma = \{\tau(h) \mid h \in H\}$. Take $G = M \cong T \rtimes O_2$ as above. The sets*

$$\tilde{R} = O_2, \qquad \tilde{T} = \bigcup_{t\in\Gamma} C_{\tilde{R}}(t)$$

are compatible.

Proof. We first note that in the present setting, for all $r \in R$, $r = r_u = r_v$. In addition, Lemma 1 says that for all $t \in T$, $r \in O_2$, $rtr^{-1} \in T$. We can therefore apply Proposition 1 to verify Condition (1) in Definition 1 for the choice of $\tilde{T}$ above. Since $\tilde{R}$ is comprised of orthogonal operators, Condition (2) is immediately satisfied. Condition (3) requires that for every $t_a \in \tilde{T}$, the magnitude of the translation vector a must be limited so that $x + a \in v$ for any $x \in u$. We assume that every $h_a \in H$ never takes a point in u outside of v by definition. Then since $\tilde{T}$ is constructed as the union of conjugacy classes corresponding to the elements of H, every $t' \in \tilde{T}$ can be seen as a rotation and/or reflection of some point in v, and Condition (3) is satisfied. □

Proposition 3. *Assume that the input spaces $\{\mathrm{Im}(v_i)\}_{i=1}^{n-1}$ are endowed with a norm inherited from $\mathrm{Im}(v_n)$ by restriction. Then at all layers, the group of orthogonal operators O_2 is the only group of transformations to which the neural response can be invariant.*

Proof. Let R denote a group of transformations to which we would like the neural response to be invariant. If the action of a transformation $r \in R$ on elements of v_i increases the length of those elements, then Condition (2) of Definition 1 would be violated. So members of R must either decrease length or leave it unchanged. Suppose $r \in R$ decreases the length of elements on which it acts by a factor $c \in [0, 1)$, so that $\|A_r(x)\| = c\|x\|$. Condition (1) says that for every $t \in \tilde{T}$, we must be able to write $t = rt'r^{-1}$ for some $t' \in \tilde{T}$. Choose $t_v = \arg\max_{\tau\in\tilde{T}} \|A_\tau(0)\|$, the largest magnitude translation. Then $t = rt'r^{-1} \Rightarrow t' = r^{-1}t_v r = t_{r^{-1}(v)}$. But $\|A_{t'}(0)\| = c^{-1}\|v\| > \|v\| = \|A_{t_v}(0)\|$, so t' is not an element of $\tilde{T}$ and Condition (1) cannot be satisfied for this r. Therefore, we have that the action of $r \in R$ on elements of v_i, for all i, must preserve lengths. The group of transformations which preserve lengths is the orthogonal group O_2. □

Proposition 4. *Let H be the set of translations associated to an arbitrary layer of the hierarchical feature map and define the injective map $\tau : H \to T$ by $h_a \mapsto t_a$, where a is a parameter characterizing the translation. Set $\Gamma = \{\tau(h) \mid h \in H\}$. Take $G = D_n \cong T \rtimes R$, with $T = C_n = \langle t \rangle$ and $R = C_2 = \{r, 1\}$. The sets*

$$\tilde{R} = R, \qquad \tilde{T} = \Gamma \cup \Gamma^{-1}t^{-u}$$

are compatible.

Proof. Since $r_u = r_v t^{-u} = t^u r_v$, we have that for $x \in T$, $r_v x r_u^{-1} = r_v x t^u r_v^{-1}$. By construction, $T \lhd G$, so for $x \in T$, $r_v x r_v^{-1} \in T$. Since xt^u is of course an element of T, we thus have that $r_v x r_u^{-1} \in T$. In the paper, we found that:

$$x' = r_v x r_u = r_v x r_v t^{-u} = x^{-1} r_v r_v t^{-u} = x^{-1} t^{-u}. \tag{2}$$

This, together with the relation $r_u = r_u^{-1}$, gives that $x^{-1}t^{-u} = r_v x r_u = r_v x r_u^{-1}$. Therefore

$$\tilde{T} = \bigcup_{x\in\Gamma} \{x, x^{-1}t^{-u}\} = \bigcup_{x\in\Gamma} \{r_v x r_u^{-1} \mid r \in \{r, 1\}\} = \bigcup_{x\in\Gamma} \{r_v x t^u r_v^{-1} \mid r \in \tilde{R}\} = \bigcup_{x\in\Gamma'} C_{\tilde{R}}(x), \tag{3}$$

where $\Gamma' = \Gamma t^u$. Thus $\tilde{T}$ is a seen as a union of $\tilde{R}$-orbits with $r' = r'_v = r'_u, r' \in \tilde{R}$, and we can apply Proposition 1 with Γ' to confirm that Condition (1) is satisfied.

To confirm Conditions (2) and (3), one can consider permutation representations of r_u, r_v and $t \in \tilde{T}$ acting on v. Viewed as permutations, we necessarily have that $A_{r_u}(u) = u$, $A_{r_u}(v) = v$, $A_{r_v}(v) = v$ and $A_t(u) \subset v$. □

References

[1] J. Bouvrie, L. Rosasco, and T. Poggio. "On Invariance in Hierarchical Models", *Advances in Neural Information Processing Systems 22*, 2009.