Enabling Equivariance for Arbitrary Lie Groups (supplementary material)

Lachlan E. MacDonald*, Sameera Ramasinghe, Simon Lucey
Australian Institute for Machine Learning
University of Adelaide, Australia

1. homNIST

The homNIST test set[1] consists of 32 homographic perturbations of each image in the MNIST test set (each of which has been black-padded from size 28x28 to size 40x40). The homographies are randomly sampled from right Haar measure for each image in he MNIST test set, and are chosen to be sufficiently close to the identity that inspection of a random sample of 320 of the images exhibit minimal edge effects. Figure 1 gives a sample of images from our homNIST test set.

2. Proofs

Here we collect rigourous mathematical proofs of the claimed theorems. To follow the proofs of Theorems 4.2, 4.3 and 4.5, the reader will need to know some topological definitions; specifically of topological spaces and open sets, compact and locally compact Hausdorff spaces, and continuity of maps; for these we refer the reader to [2]. To follow the proof of Theorem 4.4, the reader will need to be familiar with differential forms on manifolds, their pullbacks and their integrals; for these we refer the reader to [1].

In order to prove Theorems 4.2 and 4.3, we require the following technical lemma.

Lemma 2.1. *Let X and Y be locally compact Hausdorff spaces, and let $F : X \times Y \to \mathbb{R}^n$ be a continuous function for which there exists a compact set $K \subset Y$ such that the support of F is contained in $X \times K$. If ν is a Radon measure on Y, then the map*

$$X \ni x \mapsto \int_Y F(x,y)d\nu(y) \in \mathbb{R}^n \tag{1}$$

is a continuous function on X.

Proof. The proof is almost identical to [4, Lemma 1.102]. In [4, Lemma 1.102], Y is instead a locally compact Hausdorff group, with ν being Haar measure. However the group properties are not required for the proof of continuity. The only other difference is that in [4, Lemma 1.102], F is assumed to have compact support in both the x and y variables, while we hypothesise only compact support in the y variable. This difference also does not impact the proof that Equation (1) defines a continuous map. □

[1] Available at https://www.kaggle.com/datasets/lachlanemacdonald/homnist

Theorem 4.2 in fact holds at a much greater level of generality without change to the proof.

Theorem 4.2. *Let X be any Hausdorff manifold, dx a volume form on X, and suppose that G acts smoothly on X. Let $f \in C_c(X;\mathbb{R}^K)$ and $\psi \in C_c(X,\mathbb{R}^{K\times L})$. Then the formula*

$$f * \psi(u) := \int_X f(x)\cdot\psi(u^{-1}\cdot x)\,\det(Du^{-1}(x))\,dx,$$

given for $u \in G$, defines a continuous function $f * \psi \in C(G;\mathbb{R}^L)$.

Proof. The function $G \times X \ni (u,x) \mapsto \psi(u^{-1}\cdot x)\det(Du^{-1}(x)) \in \mathbb{R}^{K\times L}$ is continuous, by continuity of ψ and by smoothness of the action of G on X. It follows then that $(u,x) \mapsto f(x)\cdot\psi(u^{-1}\cdot x)\det(Du^{-1}(x))$ is continuous on $G \times X$ with support in $G \times C$, where C is the compact support of f. Once again, Lemma 2.1 implies that $u \mapsto \int_X f(x)\cdot\psi(u^{-1}\cdot x)\det(Du^{-1}(x))dx$ is a continuous function on G. □

Theorem 4.3 follows by a similar argument to that of Theorem 4.2.

Theorem 4.3. *Define the convolution of a feature map $f \in C(G;\mathbb{R}^K)$ with a filter map $\psi \in C_c(G;\mathbb{R}^{K\times L})$ by the formula*

$$f * \psi(u) := \int_G f(v)\cdot\psi(v^{-1}u)d\mu_L(v) \tag{2}$$

$$= \int_G f(uv^{-1})\cdot\psi(v)d\mu_R(v), \tag{3}$$

where $\cdot$ denotes matrix multiplication. Then $f * \psi$ is a continuous function on G, and for all $v \in G$ one has $L_v(f) * \psi = L_v(f * \psi)$.

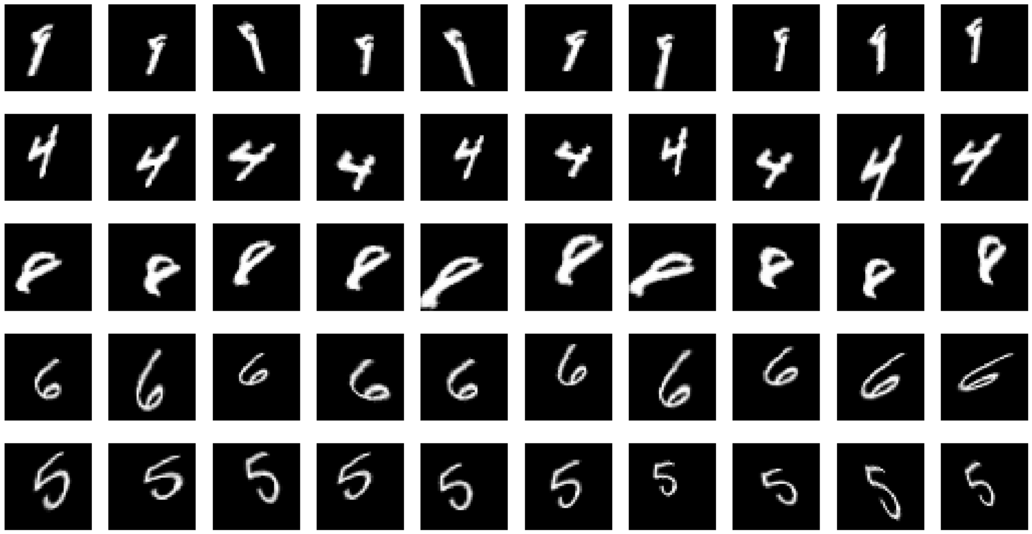

Figure 1: Samples taken from homNIST test set.

Proof. By continuity of f together with continuity of the multiplication in G, the function $(u, v) \mapsto f(uv^{-1})$ is continuous as a map $G \times G \to \mathbb{R}^K$. Let C be the compact support of ψ. We then see that $(u, v) \mapsto f(uv^{-1}) \cdot \psi(v)$ is a continuous function $G \times G \to \mathbb{R}^L$, with support contained in the set $G \times C$. Lemma 2.1 then implies that $u \mapsto \int_G f(uv^{-1}) \cdot \psi(v) d\mu_R(v)$ is continuous as a function on G. The fact that

$$\int_G f(uv^{-1}) \cdot \psi(v) d\mu_R(v) = \int_G f(w) \cdot \psi(w^{-1}u) d\mu_L(w)$$

follows from making the substitution $w = uv^{-1}$ and invoking the identity $d\mu_R(w^{-1}u) = d\mu_L(u^{-1}w)$, followed by the left-invariance $d\mu_L(u^{-1}w) = d\mu_L(w)$ of μ_L. Finally, equivariance follows immediately from Equation (3), and can be seen from Equation (2) using left-invariance of μ_L. □

Theorem 4.4 below relies on some differential geometry.

Theorem 4.4. *If f is an integrable function on G which is zero outside of a sufficiently small neighbourhood of the identity, then Haar measure $d\mu_R$ can always be chosen such that*

$$\int_G f(u) d\mu_R(u) = \int_{\mathfrak{g}} f(\exp(\xi)) \det \left(\frac{1 - e^{-ad_{-\xi}}}{ad_{-\xi}} \right) d\xi, \tag{4}$$

where $d\xi$ denotes a Euclidean volume element in the vector space $\mathfrak{g}$, and $ad_\xi : \mathfrak{g} \to \mathfrak{g}$ and $(1 - e^{-ad_\xi})/ad_\xi$ are given in the Schur-Poincaré formula for the derivative of the exponential map (Theorem 3.1).

Proof. We begin by constructing a *left* Haar measure μ_L for which an analogous formula holds, and then the result follows from the identity $d\mu_L(u) = d\mu_R(u^{-1})$. Let $d\xi$ be the standard Euclidean volume element in the vector space $\mathfrak{g}$ defined with respect to some choice of basis. Thus $d\xi = d\xi^1 \wedge \cdots \wedge d\xi^n$ where $(\xi^1, \ldots, \xi^n)$ are the coordinates defined by the basis. The cotangent multi-vector $d\xi|_0$ obtained by evaluating the form $d\xi$ at $0 \in \mathfrak{g}$ is then a nonzero volume element at the identity of G. Now the pullback formula

$$(d\mu_L)_u := L^*_{u^{-1}} d\xi|_0$$

defines a top-degree form on G, which is left-invariant since

$$\begin{aligned}(L^*_v d\mu_L)_u =& L^*_v (d\mu_L)_{vu} = L^*_v L^*_{u^{-1}v^{-1}} d\xi|_0 \\ =& (L_{u^{-1}v^{-1}} \circ L_v)^* d\xi|_0 = L^*_{u^{-1}} d\xi|_0 = (d\mu_L)_u\end{aligned}$$

for any $v \in G$. Thus $d\mu_L$ is a left Haar measure.

Now if f is an integrable function which is zero outside of a neighbourhood of the identity onto which the exponential map is a diffeomorphism, then we have

$$\int_G f(u) d\mu_L(u) = \int_{\mathfrak{g}} f(\exp(\xi))(\exp^* d\mu_L)(\xi),$$

so we must compute $\exp^* d\mu_L$. Fix $\xi \in \mathfrak{g}$, and let $\{t \mapsto \xi_i(t) = \xi + t\xi^i\}_{i=1}^n$ be the curves in $\mathfrak{g}$ through ξ pointing in the n coordinate directions. Let $p(\xi)$ be the power series $(1 - e^{-ad_\xi})/ad_\xi$. Then we compute

$$\begin{aligned}
&\big(\exp^* d\mu_L\big)_\xi \big(\xi_1'(0) \wedge \cdots \wedge \xi_n'(0)\big)\\
=&(d\mu_L)_{\exp(\xi)}\big(\exp(\xi_1)'(0) \wedge \cdots \wedge \exp(\xi_n)'(0)\big)\\
=&L^*_{\exp(-\xi)} d\xi|_0 \big(L_{\exp(\xi)} p(\xi)\xi^1 \wedge \cdots \wedge L_{\exp(\xi)} p(\xi)\xi^n\big)\\
=&d\xi|_0\big(p(\xi)\xi^1 \wedge \cdots \wedge p(\xi)\xi^n\big) = \det(p(\xi)).
\end{aligned}$$

Here, the first equality follows from the definition of the pullback, the second from Theorem 3.1, the third from the definition of $d\mu_L$ and the final from the fact that the top-degree wedge product of a linear map is equal to its determinant, together with the identity $d\xi|_0(\xi^1 \wedge \cdots \wedge \xi^n) = d\xi^1(\xi^1) \cdots d\xi^n(\xi^n) = 1$. It follows that

$$\exp^* d\mu_L(\xi) = \det\left(\frac{1 - e^{-ad_\xi}}{ad_\xi}\right) d\xi,$$

hence

$$\exp^* d\mu_R(\xi) = \det\left(\frac{1 - e^{-ad_{-\xi}}}{ad_{-\xi}}\right) d\xi$$

follows from the identities $d\mu_R(u) = d\mu_L(u^{-1})$ and $\exp(\xi)^{-1} = \exp(-\xi)$. ☐

Finally we come to Theorem 4.5. To prove it, we require the following topological result.

Lemma 2.2. *[3, Proposition 3] Let Λ be a compact topological space, and let $\{V_\lambda\}_{\lambda\in\Lambda}$ be a family of open subsets of some other topological space X. Then the intersection $\bigcap_{\lambda\in\Lambda} V_\lambda$ is an open set in X.* ☐

Theorem 4.5. *Let $f : G \to \mathbb{R}$ be a (possibly unbounded) continuous function, and $K \subset G$ a compact neighbourhood of the identity. If f takes its maximum value over K only at points in the interior of K, then there is an open neighbourhood V of the identity for which $\max_{u\in K} L_v(f)(u) = \max_{u\in K} f(u)$ for all $v \in V$.*

Proof. Let M denote the maximum value of f over K. We will prove the theorem by showing that there exist an open neighbourhood V_1 of the identity such that $\max_{u\in K} f(v^{-1}u) \leq M$ for all $v \in V_1$, and an open neighbourhood V_2 of the identity for which $\max_{u\in K} f(v^{-1}u) \geq M$ for all $v \in V_2$. Then $V := V_1 \cap V_2$ will be an open neighbourhood of the identity for which $\max_{u\in K} f(v^{-1}u) = M$ for all $v \in V$. In showing the existence of both V_1 and V_2 we will use the following observation. For each $u \in G$, let $\alpha_u : G \to G$ denote the map $v \mapsto v^{-1}u$. Then by continuity of the group operations, each α_u is continuous. Therefore, letting K° denote the interior of K (an open set), the pre-image $\alpha_u^{-1}K^\circ$ of K° under α_u is open for every $u \in G$. In particular, since $f^{-1}\{M\} \cap K$ is contained in K° by hypothesis, for all $u \in f^{-1}\{M\} \cap K$ the set $\alpha_u^{-1}K^\circ$ is an open neighbourhood of the identity.

Existence of V_1 now follows easily. For each $u \in f^{-1}\{M\} \cap K$, let $V_{1,u} := \alpha_u^{-1}K^\circ$ be the open neighbourhood of the identity considered in the previous paragraph. The set $f^{-1}\{M\} \cap K$ is a closed subset of a compact set, hence compact. Thus, by Lemma 2.2, $V_1 := \bigcap_{u\in f^{-1}\{M\}\cap K} V_{1,u}$ is an open neighbourhood of the identity such that $v^{-1}u \in K$ for all $v \in V_1$ and $u \in f^{-1}\{M\} \cap K$. It follows that $\max_{u\in K} f(v^{-1}u) \geq M$ for all $v \in V_1$.

We now come to showing the existence of V_2, which will be an intersection over $u \in K$ of open neighbourhoods $V_{2,u}$ of the identity, for which $v^{-1}u \in f^{-1}(-\infty, M]$ for all $v \in V_{2,u}$ and $u \in K$. Write K as the union $K_1 \cup K_2$, where $K_1 := K \cap f^{-1}(-\infty, M)$ and $K_2 := K \cap f^{-1}\{M\}$. For all $u \in K_1$, the continuity of α_u implies that $V_{1,u} := \alpha_u^{-1} f^{-1}(-\infty, M) \subset \alpha_u^{-1} f^{-1}(-\infty, M]$ is an open set, which contains the identity since $u \in K_1$. For $u \in K_2$, we take $V_{2,u}$ to be the open neighbourhood $\alpha_u^{-1}K^\circ \subset \alpha_u^{-1} f^{-1}(-\infty, M]$ of the identity described in the first paragraph. Now by Lemma 2.2, $V_2 := \bigcap_{u\in K} V_{2,u}$ is an open neighbourhood of the identity such that $\max_{u\in K} f(v^{-1}u) \leq M$ for all $v \in V_2$, as required. ☐

References

[1] J. M. Lee. *Introduction to Smooth Manifolds*. Number 218 in Graduate Texts in Mathematics. Springer, 2013. 1

[2] James Munkres. *Topology*. Prentice Hall, Inc, 2000. 1

[3] Leopoldo Nachbin. Compact unions of closed subsets are closed and compact intersections of open subsets are open. *Portugaliae mathematica*, 94(4):403–409, 1992. 3

[4] Dana Williams. *Crossed Products of C^*-Algebras*, volume 134 of *Mathematical Surveys and Monographs*. American Mathematical Society, 2007. 1