

A. Symmetries of the architecture

A.1. Proof of Proposition 4.1

In this appendix we provide more details on the equivariance of the Timewarp architecture. We first prove Proposition 4.1 from the main body:

Proof. Let $X(t+\tau)_{x(t)}$ denote the random variable obtained by sampling $Z \sim p(z)$ and computing $X(t+\tau) := f(Z; x(t))$. Here we subscript $X(t+\tau)_{x(t)}$ by $x(t)$ to emphasize that this is the random variable obtained when conditioning the flow on $x(t)$. We first note that the equivariance condition on the densities $p(\sigma x(t+\tau)|\sigma x(t)) = p(x(t+\tau)|x(t))$ is equivalent to the following constraint on the random variables:

$$X(t+\tau)_{\sigma x(t)} \stackrel{d}{=} \sigma X(t+\tau)_{x(t)}, \quad (20)$$

where $\stackrel{d}{=}$ denotes equality in distribution. To see this, let p_X denote the density of the random variable X . Then, in terms of densities, Equation (20) is equivalent to stating that, for all x ,

$$p_{X(t+\tau)_{\sigma x(t)}}(x) = p_{\sigma X(t+\tau)_{x(t)}}(x) \quad (21)$$

$$= p_{X(t+\tau)_{x(t)}}(\sigma^{-1}x), \quad (22)$$

where in Equation (22) we used the change-of-variables formula, along with the fact that the group actions we consider (rotations, translations, permutations) have unit absolute Jacobian determinant. Redefining $x \leftarrow \sigma x$, we get that for all x ,

$$p_{X(t+\tau)_{\sigma x(t)}}(\sigma x) = p_{X(t+\tau)_{x(t)}}(x). \quad (23)$$

Recalling the notation that $X(t+\tau)_{x(t)}$ is interpreted as the random variable obtained by conditioning the flow on $x(t)$, this can be written as

$$p(\sigma x|\sigma x(t)) = p(x|x(t)) \quad (24)$$

which is exactly the equivariance condition stated in terms of densities above. Having rephrased the equivariance condition in terms of random variables in Equation (20), the proof of Proposition 4.1 is straightforward.

$$X(t+\tau)_{\sigma x(t)} := f(Z, \sigma x(t)) \quad (25)$$

$$\stackrel{d}{=} f(\sigma Z, \sigma x(t)) \quad (26)$$

$$= \sigma f(Z, x(t)) \quad (27)$$

$$:= \sigma X(t+\tau)_{x(t)}, \quad (28)$$

where in Equation (26) we used the fact that the base distribution $p(z)$ is σ -invariant. \square

A.2. Translation equivariance via canonicalisation

We now describe the canonicalisation technique used to make our models translation equivariant. Let $q(x^p(t+\tau), x^v(t+\tau)|x^p(t))$ be an arbitrary conditional density model, which is not necessarily translation equivariant. We can make it translation equivariant in the following way. Let $\overline{x^p}$ denote the average position of the atoms,

$$\overline{x^p} := \frac{1}{N} \sum_{i=1}^N x_i^p. \quad (29)$$

Then we define

$$p(x^p(t+\tau), x^v(t+\tau)|x^p(t)) := q(x^p(t+\tau) - \overline{x^p(t)}, x^v(t+\tau)|x^p(t) - \overline{x^p(t)}) \quad (30)$$

where the subtraction of $\overline{x^p(t)}$ is broadcasted over all atoms. We now consider the effect of translating both $x^p(t)$ and $x^p(t+\tau)$ by the same amount. Let a be a translation vector in \mathbb{R}^3 . Then

$$p(x^p(t+\tau) + a, x^v(t+\tau)|x^p(t) + a) = q(x^p(t+\tau) + a - \overline{(x^p(t) + a)}, x^v(t+\tau)|x^p(t) + a - \overline{(x^p(t) + a)}) \quad (31)$$

$$= q(x^p(t+\tau) + a - \overline{x^p(t)} - a, x^v(t+\tau)|x^p(t) + a - \overline{x^p(t)} - a) \quad (32)$$

$$= q(x^p(t+\tau) - \overline{x^p(t)}, x^v(t+\tau)|x^p(t) - \overline{x^p(t)}) \quad (33)$$

$$= p(x^p(t+\tau), x^v(t+\tau)|x^p(t)). \quad (34)$$