

# Shape normalization for keypoint data

Caleb Weinreb and Kai Fox

May 2023

## 1 Modeling framework

The goal of shape normalization is to map keypoint observations from many animals into a standardized pose space where the effects of body morphology have been removed. Naively, one might hope to normalize using simple summary statistics like average nose-to-tail distance. But these statistics are problematic because they conflate body shape with behavior. For example, larger average nose-to-tail distance could reflect a larger body, or simply a higher frequency of stretched-out poses. Thus the central challenge of shape normalization is to isolate the effects of behavior and body shape so that the latter can be cleanly removed. To that end, we propose a generative model that factors behavior and morphology into separate terms that are disentangled during fitting.

### 1.1 Input data

We consider a dataset of poses  $\{y_{n,t}\} \subset \mathbb{R}^{KD}$  measured from animals  $n = 1, \dots, N$  at timepoints  $t = 1, \dots, T_n$  (the timepoints need not be sequential), where each pose contains coordinates of  $K$  keypoints in  $D$  dimensions. We assume that the keypoints are already centered and rotationally aligned, and that differences in body size have already been minimized as much as possible through uniform scaling of each animal.

### 1.2 Generative model

We model each pose  $y_{n,t}$  as the animal-specific realization of a standardized pose  $x_{n,t}$  that is sampled from a Gaussian mixture. The mixture components are shared across animals, but the mixture weights are animal-specific. Formally, the model is

$y_{n,t} = f(x_{n,t}, \phi_n)$	(observed pose given standardized pose)
$x_{n,t} \sim \mathcal{N}(m_{z_{n,t}}, Q_{z_{n,t}})$	(standardized pose given mixture component)
$z_{n,t} \sim \text{Cat}(\pi_n)$	(mixture component given animal-specific weights)
$m_z \sim \mathcal{N}(m_0, \lambda_0^{-1} Q_z)$	(prior on mixture means given covariances)
$Q_z \sim \text{Wishart}^{-1}(W_0, \nu_0)$	(prior on mixture covariances)
$\pi_n \sim \text{Dir}(\beta)$	(prior on animal-specific mixture weights)
$\beta \sim \text{Dir}(\gamma, \dots, \gamma)$	(prior on mixture weight hyperparameters)

where  $f$  is a morph function that maps standardized poses  $x_{n,t}$  to animal-specific poses  $y_{n,t}$  and  $\phi_n$  is a parameter vector that captures the unique shape of animal  $n$ . Several forms for  $f$  (and priors on  $\phi_n$ ) are described in Section XXX and tested in the paper. In general,  $f$  should be invertible in its first argument and differentiable in its second. Abusing notation, we will use  $f^{-1}$  to denote the inverse of  $f$  in its first argument.

### 1.3 Expectation maximization

The model parameters  $\theta = (\phi, m, Q, \pi, \beta)$  are fit using expectation maximization (EM), which aims to maximize the expected log likelihood  $\ell(\theta) = \mathbb{E} \log P(y, z | \theta)$ . EM alternates between two steps.

**E-step** Given current parameter estimates  $\theta^* = (\phi^*, m^*, Q^*, \pi^*, \beta^*)$ , calculate the posterior  $q(z) = P(z | y, \theta^*)$ , as follows.

$$q(z) = \prod_{n,t} P(z_{n,t} | y_{n,t}, \phi^*, m^*, Q^*, \pi^*) \quad (1)$$

$$\propto \prod_{n,t} \mathcal{N}(f^{-1}(y_{n,t}, \phi_n^*) | m_{z_{n,t}^*}^*, Q_{z_{n,t}^*}^*) \cdot \pi_{n,z_{n,t}^*}^* \quad (2)$$

**M-step** Given  $q(z)$  from the E-step, obtain new parameter estimates by optimizing the following objective with respect to  $\theta = (\phi, m, Q, \pi, \beta)$  (using gradient ascent).

$$A(\theta; \theta^*) = \sum_{n,t} \mathbb{E}_q \log P(y_{n,t}, z_{n,t} | \theta) + \log P(\theta) \quad (3)$$

$$= \sum_{n,t} q(z_{n,t}) \log [\mathcal{N}(f^{-1}(y_{n,t}, \phi_n) | m_{z_{n,t}}^*, Q_{z_{n,t}}^*) \cdot \det(J) \cdot \pi_{n,z_{n,t}}] + \log P(\theta) \quad (4)$$

where  $J$  is the Jacobian of  $f^{-1}$  (in its first argument) evaluated at  $y_{n,t}$ .

### 1.4 Initialization

Initialization of the morph parameters  $\phi$  is described in section XXX. Given morph parameters, we fit a standard Gaussian mixture model to the points  $x_{n,t} = f^{-1}(y_{n,t}, \phi_n)$  (e.g. with sklearn), yielding mixture components  $(m_z, Q_z), z = 1, \dots, L$  and cluster weights  $\lambda_{n,t,z}$  for each data point. Using the weights, we initialize  $\pi$  and  $\beta$  as follows.

$$\pi_{n,z} = \frac{1}{T_n} \sum_{t=1}^{T_n} \lambda_{n,t,z}, \quad \beta_z = \frac{1}{N} \sum_{n=1}^N \pi_{n,z} \quad (5)$$

## 2 Morph models

XXX

### 3 E-step

In the E-step we seek to calculate  $q_{n,t}(z) = P(z | y_{n,t}, \theta^*)$  for a given keypoint observation  $y_{n,t}$  and estimated parameters  $\theta^*$ . This is usually done using Bayes' rule

$$q_{n,t}(z) = P(z | y_{n,t}, \theta^*) = \frac{P(y_{n,t} | z, \theta^*) P(z | \theta^*)}{\sum_{z'} P(y_{n,t} | z', \theta^*) P(z' | \theta^*)} \quad (6)$$

Each probability above is given directly by the generative model (Sec ??), and so may be expanded as

$$q_{n,t}(z) = \frac{\mathcal{N}(f^{-1}(y_{n,t}, \phi_n^*) | m_z^*, Q_z^*) \cdot \pi_{n,z}^*}{\sum_{z'} \mathcal{N}(f^{-1}(y_{n,t}, \phi_n^*) | m_{z'}^*, Q_{z'}^*) \cdot \pi_{n,z'}^*} \quad (7)$$

which yields the proportionality result stated in Eq. ?? with the additional specification of normalization so that  $q_{n,t}$  is a probability distribution in  $z$ .