Shape normalization for keypoint data

Caleb Weinreb and Kai Fox

May 2023

1 Modeling framework

The goal of shape normalization is to map animal poses into a common (possibly low-dimensional) space where the effect of body shape has been removed. Naively, one might try to normalize using simple summary statistics like average nose-to-tail distance. The problem is, such statistics conflate body shape with behavior. For example, large average nose-to-tail distance could reflect a larger body, or simple 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 factored out. To that end, we propose a generative model that embodies morphology and behavior in two separate sets of parameters and then fits them simultaneously to a dataset of measured poses from a population of animals.

1.1 Generative model

Formally, suppose that we measure the poses $\{y_{n,t}\}$ of animals $n = 1, ..., N$ at timepoints $t = 1, ..., T_n$ (the timepoints need not be sequential), where each pose $y_{n,t} \in \mathbb{R}^{KD}$ is a vector of K keypoints in egocentric coordinates. We assume that each measured pose $y_{n,t}$ is the animal-specific realization of an idealized pose $x_{n,t} \in \mathbb{R}^M$ ($M \leq KD$) as follows:

$$y_{n,t} = f_n(x_{n,t}), \ x_{n,t} \sim P_n \tag{1}$$

where $\xi \sim \mathcal{N}(0, R)$ is measurement noise, P_n is a distribution over idealized pose states, and f_n is a *morph* function that maps idealized poses to observed poses animal n. Our goal in fitting these parameters is to explain each animal's pose distribution through a combination of morphological and behavioral variation, embodied in the variables f_n and P_n repsectively. To encourage this partition, we assume that the f_n's and P_n's have a specific functional form defined below.

Morph functions: We assume each f_n is an affine map with coefficients that depend on a parameter vector ϕ. That is:

$$f_n(x) = C(\phi_n)x + d(\phi_n) \tag{2}$$

Note that the affine assumption does not limit the morph functions to simple scalings and shears in D-dimensional keypoint space, but rather allows for complex, pose-dependent shifts of the keypoint coordinates. We do not specify the form of the mapping $\phi \mapsto (C(\phi), d(\phi))$ at this point. Several forms are considered in section XXX.

Pose distributon: Animals of the same species typically share the same basic repertoire of poses and differ mainly in how often each pose is realized. To capture this structure, we assume that the distributions P_n are Gaussian mixtures, where the mixture modes $\mathcal{N}(m_z, Q_z)$ are shared across animals. The mixture weights π_n differ across animals but share a common prior β. Formally:

$$x_{n,t} \sim \mathcal{N}(m_{z_{n,t}}, Q_{z_{n,t}}), \qquad z_{n,t} \sim \text{Cat}(\pi_n), \qquad \pi_n \sim \text{Dir}(\beta), \qquad \beta \sim \text{Dir}(\gamma, \ldots, \gamma)$$

1.2 Expectation maximization

The model parameters $\theta = (\phi, m, Q, \pi, \beta)$ can be fit using expectation maximization (EM), which aims to maximize the expected log likelihood $\ell(\theta) = \mathbb{E}_{x,z} \log P(y, x, z \mid \theta)$ by maintaining an auxiliary distribution $q(x, z \mid y, \theta^*)$ that approximates the posterior $P(x, z \mid y, \theta^*)$ for the current parameter estimate θ^*. During each iteration of EM, the auxiliary distribution q is updated (E-step), and then new parameters θ^*_{new} are calculated (M-step) by maximizing objective

$$A(\theta; \theta^*) = \mathbb{E}_{q(x,z|y,\theta^*)} \log P(y, x, z|\theta) + \log P(\theta) \tag{3}$$

The theoretical basis for EM (see Murphy 11.4.7 [cite]) is that this procedure guarantees monotonic increases in the expected log likelihood, i.e. $\ell(\theta^*_{\text{new}}) \geq \ell(\theta^*)$. In our case, since the model likelihood factors over animals and timepoints, the objective can be rewritten as follows.

$$A(\theta; \theta^*) = \sum_{n,t,z} \int_x q_{n,t}(x, z) \, \log P(y_{n,t}, x, z \mid \theta) \, dx + \log P(\theta) \tag{4}$$

Here we derive analytic expressions for the auxiliary distrubutions as well as the above integral, and perform the M-step using gradient ascent.

1.3 Responsibilities of morph and pose space models in EM

The majority of the calculations in both the E- and M-steps can be performed in terms of the affine map (C_n, d_n) and therefore may be posed as responsibilities of the pose space model, given the affine map produced by the morph model as input. The only responsibilities of the morph model are to calculate these matrices and vectors, evaluate priors and initializations for the parameters, and to provide convenience functions such as finding maximum likelihood estimates of latents given observations and parameters.

The requirements of the pose space model on the other hand are to calculate priors, initializations, and convenience functions as the morph model does, but also to evaluate the objective function A of the maximization routine in Eq. 4. We now expand upon the framework for analytically integrating our problematic integral. In doing so we enumerate both the necessary arithmetic results to derive for the M-step of each pose space model, and the computations to be performed in their E-steps.

The theoretical guarantees of EM are achieved by taking the latent distribution to be the log likelihood of the latents given the observations and current parameter estimates θ^*, which is usually calculated using Bayes' rule

$$q_n^t(x, z) = P(x, z \mid y_n^t, \theta^*) = \frac{P(y_n^t, x \mid z, \theta^*) P(z \mid \theta^*)}{P(y_n^t \mid \theta^*)} \tag{5}$$

For integration in x, this fraction breaks down as a normalizer times $P(y_n^t, x \mid z, \theta^*)$. As shown below, that probability the product the normal PDF around $f(x)$ giving the probability of the noisy observation y_n^t and the probability of the pose state x given by F. It will be the central task of the M-step to rewrite that product in turn as a constant in x times an auxiliary probability distribution s to be integrated against the data likelihood $\log P(y_n^t, x, z \mid \theta)$:

$$P(y_n^t, x \mid z, \theta^*) = P(y_n^t \mid x, \theta^*) P(x \mid z, \theta^*) \tag{6}$$

$$= \mathcal{N}(y_n^t \mid C_n(\phi^*) \, x + d_n(\phi^*), R^*) F(x \mid z, \psi^*) \tag{7}$$

$$:= K_{y_n^t, z, \theta^*} \cdot s(x; y_n^t, z, \theta^*) \tag{8}$$

We expand upon the requirement that $s(x)$ be a PDF in which expectations of the log probability terms arising from the data likelihood are known below in Note that for brevity we write s, K as functions of θ, however by Eq. 7 they are functions only of $(\psi^*, R^*, C_n(\phi^*), d_n(\phi^*))$ and therefore may be evaluated by the morph model operating only on the affine transform output by an arbitrary pose space model.

We find that these constants K and expectations in s, together with the finite array of probabilities for the discrete latents $G_n(z \mid \gamma)$, are sufficient to express the objective function $A(\theta; \theta^*)$. The remaining term in the numerator of q_n^t in Eq. 5, $P(z \mid \theta^*)$, is simply $G(z \mid \gamma^*)$. Additionally, marginalizing Eq. 8 over x shows that the constants K are in fact the probabilities $P(y_n^t \mid z, \theta^*)$, so the denominator $P(y_n^t \mid \theta^*)$ of q_n^t in Eq. 5 may be expanded by the law of total probability as $\sum_z K_{y_n^t, z, \theta^*} G(z \mid \gamma^*)$. The objective function as defined in Eq. ?? may therefore be written

$$A(\theta; \theta^*) = \sum_{t,n} \sum_z \frac{K_{y_n^t, z, \theta^*} G(z \mid \gamma^*)}{\sum_{z'} K_{y_n^t, z', \theta^*} G(z' \mid \gamma^*)} \, \mathbb{E}_{s(x; y_n^t, z, \theta^*)} \left[\log P(y_n^t, x, z \mid \theta) \right] + \log P(\theta) \qquad (9)$$

Using Eq. 9, we may state explicitly the arithmetic work to be done in deriving a pose space model and the computations that a pose space model must perform during an EM iteration in our modeling framework:

Derivation: A pose space model must define (F, G), and rewrite the product of F with a Gaussian probability as a constant K times a PDF $s(x)$, as in Eq. 8. Closed form integrals of s against the two terms outlined in Sec. 1.4 must be derived.

E step: A pose space model must, during the E-step, calculate the constants $K_{y_n^t, z, \theta^*}$ as a function of the estimated parameters θ^* resulting in an $\mathbb{R}^{|y| \times |z|}$ matrix, where $|y|$ is the number of observed samples and $|z|$ is the cardinality of the discrete latent "phoneme" space. The function s will be defined parametrically, and during the E-step a pose space model must also calculate the $\mathbb{R}^{|y| \times |z|}$ matrix of parameters as a function of θ^*.

M step: A pose space model will use the derived closed form expectations as functions of θ and s (parameterized as the results of the E-step) alongside the constants K calculated in the E-step to compute the objective function Eq. 9 during the M-step. General model-agnostic methods will perform gradient-based maximization based on this output.

1.4 Requirements for the auxiliary distribution of a pose model

It is crucial for each pose space model to formulate an auxiliary PDF s for which the expectations in the rewritten objective function (Eq. 9) have a closed form expression. To elaborate on this constraint, we expand the log probability of the data using the structure of our generative model (Eq. ??):

$$\log P(y_n^t, x, z \mid \theta) = \log P(y_n^t \mid x, \theta) + \log P(x \mid z, \theta) + \log P(z \mid \theta) \qquad (10)$$
$$= \log \mathcal{N}(y_n^t \mid C_n(\phi) \, x + d_n(\phi), R) + \log F(x \mid z, \psi) + \log G(z \mid \gamma) \qquad (11)$$

Of these, only the first two are functions of x and thereby impose a requirement on $s(x)$ of having a known closed form expectation.

It will in general be necessary when finding the closed form expectations to recognize that the first term may be written as a normal PDF in x by transforming the normal distribution of y around $f(x)$ to a scaled normal distribution of x around $f^{-1}(y)$. Let $||a - b||_A$ is the Mahalanobis distance

of a, b according to covariance A. Then, dropping the dependence of C_n, d_n on ϕ for readability,

$$\log \mathcal{N}(y \mid C_n\, x + d_n, R) = \log\left(\frac{Z_R}{Z_{C_n^{-1}RC_n^{T^{-1}}}}\mathcal{N}\left(x \mid C_n^{-1}\left(y - d_n\right),\ C_n^{-1}R\,C_n^{T^{-1}}\right)\right) \tag{12}$$

$$= \log Z_R - \frac{1}{2}\left|\left|x - C_n^{-1}(y - d_n)\right|\right|^2_{C_n^{-1}R\,C_n^{T-1}} \tag{13}$$

The result is an unnormalized Gaussian log probability to be integrated against $s(x)$ and a normalizer term that may be computed independent of the morph model. For completeness, we now explicate a proof general inversion of an affine-transformed Gaussian PDF:

Proof. For vectors $x \in \mathbb{R}^M$ and $y \in \mathbb{R}^N$ and matrices $A \in \mathbb{R}^{N \times M}, B \in \mathbb{R}^{N \times N}$, we may expand a normal PDF as a normalizer times a squared Mahalanobis distance:

$$\mathcal{N}(y \mid Ax, B) = Z_B \exp\left\{-\frac{1}{2}\left|\left|y - Ax\right|\right|^2_B\right\} \tag{14}$$

$$= Z_B \exp\left\{-\frac{1}{2}(y - Ax)^T B^{-1}(y - Ax)\right\} \tag{15}$$

$$= Z_B \exp\left\{-\frac{1}{2}(A^{-1}y - x)^T A^T B^{-1} A(A^{-1}y - x)\right\} \tag{16}$$

$$= Z_B \exp\left\{-\frac{1}{2}\left|\left|A^{-1}y - x\right|\right|^2_{A^{-1}BA^{-1T}}\right\} \tag{17}$$

$$= \frac{Z_B}{Z_{A^{-1}BA^{-1T}}}\mathcal{N}\left(x \mid A^{-1}y,\ A^{-1}BA^{-1T}\right) \tag{18}$$

Additionally, for x, y, z in $\mathbb{R}^N$ and $B \in \mathbb{R}^{N \times N}$, $\mathcal{N}(x - z \mid y, B) = \mathcal{N}(x \mid y + z, B)$. Applying these two identities yields Eq. 13. $\qquad\square$

To summarize, during the derivation of a pose space model, expectations in the random variable $x \sim s$ of the following terms must be derived as closed-form expressions. Note that while s is defined in terms of estimated parameters θ^*, the following terms are functions of test parameters θ during gradient based optimization of A.

- The Gaussian log-PDF centered at morphed latent poses $f(x) = C_n x + d_n$ with observation covariance R evaluated at observations y_n^t. This PDF breaks down as in Eq. 13 into a quadratic form and a Gaussian normalizer $\log Z_R$ that is constant in x.

- The log-PDF of the pose space distribution $F(x \mid z, \psi)$ at a latent pose x under given discrete pose state z and test parameters ψ.

2 Morph models

2.1 Scalar morph

Before introducing spatially nonlinear morphs, we begin by working with a simple uniform scale around an affine offset. Letting the morph model be parameterized by scalars α_n and offsets $\mu_n \in \mathbb{R}^N$, we take $C_n(\phi) = e^{\alpha_n}I_{KD}$, and $d_n(\phi) = \mu_n$. Using this scalar morph the latent pose space dimension cannot compress keypoint space, so $M = KD$.

Priors As will be standard for most morph models, we define a prior to remove ambiguity both in the offsets μ_n and the scale parameters α_n. In particular, if $\bar{\mu}(\phi), \bar{\alpha}(\phi)$ are the average offset

$\mathbb{E}_n[\mu_n(\phi)]$ and log-scale, respectively, then the prior on ϕ for the scalar morph is

$$L_{\text{morph}}(\phi) = \log \mathcal{N}\left(\bar{\mu}(\phi) \mid 0, I_{KD}\right) + \log \mathcal{N}\left(\bar{\alpha}(\phi) \mid 0, I_1\right) \tag{19}$$

$$= -\frac{1}{2}\left\|\bar{\mu}(\phi)\right\|_2^2 - \frac{1}{2}\bar{\alpha}(\phi)^2 \tag{20}$$

Initialization   We initialize offsets to the subject-wise mean in keypoint space $\mu_n = \mathbb{E}_t[y_n^t]$ and log-scale factors to the MLE of standard deviation for a spherical Gaussian centered at μ_n given the subject's keypoint data, $\alpha_n = \frac{1}{2}\log \mathbb{E}_t[||y_n^t - \mu_n||_2^2]$

2.2   Affine modal alignment

The first (dare I say ...only?) spatially non-rigid morph we will consider is an affine linear transform which aligns a limited number of dimensions around a centroid. We first give the intuition for this model, then define it and specify priors and an initialization routine.

As in the scalar morph model, we make the assumption that all latent poses for a subject are related to their keypoints by the same offset and scaling, i.e. that there is a certain homologous centroid pose that varies per-animal and that each animal is characterized by a scale of variations from that centroid. In addition, we consider that there might be variations from the centroid that differ by subject but should also be considered homologous — for example large mice might turn more sharply to follow the same path, or that the limbs of older mice might take flatter trajectories when walking if their bodies are held lower to the ground. We therefore allow variation of a certain number, L, of posture *modes* to be mapped differently from latent pose space to each animal's body. We may represent a pose as a magnitude of variation from the centroid along each of these dimensions (as is done in PCA) along with an additional $(KD - L)$-dimensional component describing the remaining variation, which we leave unaltered when mapping from pose space to keypoint space.

Let L be an integer d.o.f. hyperparameter giving the number of pose modes to be aligned. In this case, we define the morph model in terms of the following parameters

- $\bar{\alpha} \in \mathbb{R}$ - log of the subject-wise uniform scale factor $\alpha = e^{\bar{\alpha}}$ to be applied to all dimensions of pose space

- $U \in \mathbb{R}^{KD \times L}$ - matrix whose columns give the population-wide modes in pose space that may be morphed differently per-animal. *In the special case where the columns of U are the first L principal components and are considered a hyperparameter instead of being learned, we call this morph a "PC modal alignment".*

- $\hat{U}_n \in \mathbb{R}^{KD \times L}$ - matrix whose columns, $\hat{u}_{n,l}$, give subject-wise updates to the morph dimensions.

- $\mu_n \in \mathbb{R}^{KD}$ - center of the affine transform, so that $d_n(\phi) = \mu_n$.

Let $\bar{U}$ be the projection on to the orthogonal complement of U, which will be used as a pass-through for those dimensions of pose space that should be effected only by uniform scaling. Using these parameters, the modal alignment morph is the affine map given by the matrix $C_n(\phi)$ and the vector $d_n(\phi)$:

$$C_n(\phi) = \alpha_n \left[\bar{U} + \left(U + \hat{U}_n\right)U^+\right] \qquad \text{with } \phi = (U, \bar{\alpha}_n, \hat{U}_n, \mu_n, \ n \in 1...N) \tag{21}$$

$$d(\phi_n) = \mu_n \qquad\qquad \alpha_n = \exp\left(\bar{\alpha}_n\right) \tag{22}$$

$$\hat{u}_{n,l} = (\hat{U}_n)_{:,l} \sim \mathcal{N}(0, \upsilon^2 I_{KD}) \tag{23}$$

To understand the action of the morph, consider the $L = 2$ case in which the columns of U, which we call u_1 and u_2, are orthonormal. Let $\hat{u}_1, \hat{u}_2$ be the updates to the morph dimensions for subject n. Then the action of $C(\phi_n)$ on $\mathrm{span}(u_1, u_2)$ - the subspace spanned by the morph dimensions - is a perturbation of the symmetric matrix with SVD $U(\alpha I_{KD})U^T$, namely

$$C(\phi_n) : x \in \mathrm{span(U)} \mapsto \begin{bmatrix} | & | \\ u_1 + \hat{u}_1 & u_2 + \hat{u}_2 \\ | & | \end{bmatrix} \begin{bmatrix} \alpha & \\ & \alpha \end{bmatrix} \begin{bmatrix} - & u_1^T & - \\ - & u_2^T & - \end{bmatrix} x \tag{24}$$

The action of $C(\phi_n)$ on the orthogonal complement of U should then be the identity. Removing the requirement of orthogonal morph dimensions, and considering vectors outside the span of those morph dimensions, we have the general expression:

$$C(\phi_n) = \alpha \left[\bar{U} + \left(U + \hat{U}_n \right) \mathrm{Diag}\left(s_n \right) U^+ \right] \tag{25}$$

with U^+ being the Moore-Penrose pseudoinverse of U.

Priors We inherit the normal priors on the average offset and log-scale to remove ambiguity in μ_n and α_n as in Eq. 20. In addition, we include a Gaussian prior on the mode adjustments, the columns of the matrix $\hat{U}_n$ which we call $u_{n,l}$, after scaling by the norm of the corresponding mode u_l, the l-th column of the matrix U, according to a scale hyperparameter v. Finally, we remove symmetry in $||u_l||$ and arrive at the following morph parameter log likelihood function:

$$L_{\mathrm{morph}}(\phi) = \log \mathcal{N} \left(\bar{\mu}(\phi) \mid 0, I_{KD} \right) + \log \mathcal{N} \left(\bar{\alpha}(\phi) \mid 0, I_1 \right) \tag{26}$$

$$+ \sum_{n,l=1}^{N,L} \log \mathcal{N}(u_{n,l}, 0, v^2 I_{KD}) + \sum_{l=1}^{L} \log \mathcal{N}(\log ||u_l||, 0, 1) \tag{27}$$

$$= -\frac{1}{2} ||\bar{\mu}(\phi)||_2^2 - \frac{1}{2} \bar{\alpha}(\phi)^2 - \frac{1}{2v^2 KD} \sum_{n,l=1}^{N,L} ||u_{n,l}||_2^2 ||u_l||_2^{-2} - \frac{1}{2} \sum_{l=0}^{1} \log^2 ||u_l|| \tag{28}$$

3 Pose space models

3.1 Gaussian mixture

In the Gaussian mixture model, our discrete latent "phoneme" state is categorical over L values (a hyperparameter) which we term components as in a standard GMM. The discrete distribution weights subject to a heirarchical Dirichlet prior with hyperparameters for usage uniformity across components, $\hat{\beta}$, component usage uniformity across animals $\hat{\pi}$, and a saturation level for logits $\pi_{\max}$. Our continuous pose state is normal, with mean and covariance conditional on the discrete state.

$$G \left(z \mid \gamma \right) = \mathrm{Cat}(\pi_n) \qquad \text{with } \gamma = (\bar{\beta} \in \mathbb{R}^L, \bar{\pi}_n \in \mathbb{R}^L, \ n \in 1...N)$$

$$\beta = \mathrm{softmax}(\bar{\beta}), \ \beta \sim \mathrm{Dir}(\hat{\beta})$$

$$\pi_n = \mathrm{softmax}(\pi_{\max} \tanh(\bar{\pi}_n / \pi_{\max})), \ \pi_n \sim \mathrm{Dir}(\hat{\pi} \, \gamma)$$

$$F(x \mid z, \psi) = \mathcal{N}(m_z, Q_z) \qquad \text{with } \psi = ((m_z, Q_z) \in \mathbb{R}^M \times \mathbb{R}^{M \times M}, \ z \in 1...L)$$

3.1.1 Auxiliary PDF and constants

Our first task is to rewrite the product of F and a normal PDF in y_n^t, from Eq. 7, as a constant K times a PDF $s(x)$. Since F is itself Gaussian in this case, we may apply the same transformation as in 13 to arrive at a product of two Gaussians in x, which is proportional to another normal PDF in x. The following proposition specifies the proportionality constant.

Proposition 3.1. *The product normal PDFs evaluated at a point, $\mathcal{N}(x \mid a, A)$ and $\mathcal{N}(x \mid b, B)$, is proportional to $\mathcal{N}(x \mid c, C)$ with and $C = A\,(A + B)^{-1}\,B$ and $c = CA^{-1}a + CB^{-1}b$. Moreover, if $Z_\Sigma = (2\pi)^{-D/2}\,|\Sigma|^{-1/2}$ is the usual Gaussian normalization factor for covariance matrix Σ, then equality is achieved using the following proportionality constant:*

$$
\mathcal{N}(x \mid a, A)\mathcal{N}(x \mid b, B) = \frac{Z_A Z_B}{Z_C} \exp\left\{ -\frac{1}{2} \left(a^T A^{-1} a + b^T B^{-1} b - c^T C^{-1} c \right) \right\} \mathcal{N}(x \mid c, C).
\tag{29}
$$

Proof. The proportionality result is standard, so we leave that proof to the reader and use the result to derive our proportionality constant. Let $||x - a||_A^2 = (x - a)A^{-1}(x - a)^T$ be the squared Mahalanobis distance between x and a under covariance A, so that $\mathcal{N}(x \mid a, A) = Z_A \exp(-\frac{1}{2}\,||x - a||_A^2)$. The normalizing constant may therefore be written

$$
\frac{\mathcal{N}(x \mid a, A)\mathcal{N}(x \mid b, B)}{\mathcal{N}(x \mid c, C)} = \frac{Z_A Z_B}{Z_C} \exp\left\{ -\frac{1}{2} \left(||x - a||_A^2 + ||x - b||_B^2 - ||x - c||_C^2 \right) \right\}
\tag{30}
$$

Expanding the Mahalanobis distances, we see that the remaining terms are precisely those which to not vary in x (which is natural given the proportionality result) yielding the desired result. □

The terms $K_{y_n^t, z, \theta_{C,n}}$ and $s(x; y_n^t, z, \theta^*)$ for the Gaussian mixture pose space model may be calculated using Proposition 3.1 by inverting the affine transform of x in Eq. 7 and applying it to y_n^t, as in 13. The mean and covariances of the constituent normals are then given in the language of the proposition by $a = C_n^{-1}(y + d_n)$, $A = C_n^{-1} R\, C_n^{T^{-1}}$, $b = m_z$, and $B = Q_z$, with dependence of C_n, d_n on ϕ^* suppressed for readability, leaving s to be the normal PDF $N(x \mid c, C)$ and K to be the proportionality constant in Eq. 29.

3.1.2 Expectations of log probabilities

It remains to be seen that s can be analytically integrated against the requisite log probability terms: the quadratic form from Eq. 13 and $\log F$, which is also a quadratic form here. To calculate these integrals, we will use the following proposition.

Proposition 3.2. *The expectation of a quadratic form in a normal variable is the sum of a Mahalanobis distance and a trace:*

$$
\mathbb{E}_{x \sim \mathcal{N}(a, B)} \left[(x - c)^T D^{-1}(x - c) \right] = ||a - c||_D^2 + \mathrm{Tr}\left[BD^{-1} \right]
\tag{31}
$$

The proof is left to the reader, but is achieved by wrapping the whole expectation in a trace and cycling its arguments to arrive at an outer product $(x - c)(x - c)^T$. We may now enumerate the log probability terms that are to be calculated by the Gaussian mixture pose space model:

- The expectation in s of the quadratic form from Eq. 13 is given by the proposition, scaled by $-\frac{1}{2}$

- The expectation in s of $\log F$ splits into two terms. First, we have the expectation of the quadratic form $-\frac{1}{2}\,||x - m_z||_{Q_z}^2$, which is again given by the scaled result of the composition. Additionally, there is a term $\log Z_{Q_z} \propto -\frac{1}{2}\log |Q_z|$ which is constant in x and so is its own expectation.

Priors   We impose a heirarchical Dirichlet prior on the component weights π_n, but we do not constrain the parameters of the component distributions m_z, Q_z. This results in the following log-prior for the Gaussian mixture pose model:

$$L_{\mathrm{morph}}(\gamma) = \sum_n \log \mathrm{Dir}(\pi_n \mid \hat{\pi}\,\beta) + \log \mathrm{Dir}(\beta \mid \hat{\beta}/L) \tag{32}$$

Note that the hyperparameters $\hat{\beta}$, $\hat{\pi}$ define the variation tolerable in the expected weight of each component and the variation tolerable from that mean respectively. In particular, $\mathbb{E}[\beta]$ is the discrete uniform over L choices, with $\mathrm{Var}[\beta_l]$ inversely proportional to $\hat{\beta} + 1$. Moreover, $\mathbb{E}[\pi_n] = \beta$ with $\mathrm{Var}[\pi_{n,l}]$ inversely proportional to $\hat{\pi} + 1$.

Initialization   We initialize the Gaussian mixture pose space model based on a standard Gaussian mixture model (GMM) fit to pose space data from a reference subject $\hat{n}$. Specifically, given parameters for a morph model, ϕ, we form the dataset of pose space points $\{C_{\hat{n}}(\phi)y_n^t + d_{\hat{n}}(\phi)\}_t$ and fit a GMM in L components, yielding means $\hat{m}_z$, covariances $\hat{Q}_z$ and cluster weights $\hat{\pi}_z$. From these we may construct initialization parameters π and ψ as follows:

- Component means and covariances of the pose space model are directly inherited from the reference-subject GMM. That is, $m_z = \hat{m}_z$ and $Q_z = \hat{Q}_z$.

- The heirarchical Dirichlet prior is initialized so that $\pi_{n,z} = \hat{\pi}_z$ and so that the initial π_n's are the mean of their generating distribution, i.e. we take the component weight logits as $\bar{\pi}_{n,z} = \log \hat{\pi}_z$, and we take $\bar{\beta}_z = \log \hat{\pi}_z$ so that $\mathbb{E}[\mathrm{Dir}(\hat{\pi}\beta)] = \pi_n$.