Analysis and synthesis of 3D shape families via deep-learned generative models of surfaces Supplementary Material

1 Mean-field inference equations

According to the mean-field approximation theory [1], given a probability distribution P defined over a set of variables $X_1, X_2, ..., X_V$, we can approximate it with a simpler distribution Q, expressed as a product of individual distributions over each variable, such that the KL-divergence of P from Q is minimized:

$$KL(Q \mid\mid P) = \sum_{X_1} \sum_{X_2} ... \sum_{X_V} Q(X_1, X_2, ..., X_V) \cdot \ln \frac{Q(X_1, X_2, ..., X_V)}{P(X_1, X_2, ..., X_V)}$$

In the case of continuous variables, the above sums are replaced with integrals over their value space. Suppose that the original distribution P is defined as a product of factors:

$$P(X_1, X_2, ..., X_V) = \frac{1}{Z} \prod_{s=1...S} \phi_s(\mathbf{D}_s)$$

where $\mathbf{D}_s$ is a subset of the random variables (called scope) for each factor s in the distribution P, and Z is a normalization constant.

Minimizing the KL-divergence of P from Q yields the following mean-field updates for each variable X_v ($v = 1...V$):

$$Q(X_v) = \frac{1}{Z_v} \exp \left\{ \sum_s \sum_{\mathbf{D}_s - \{X_v\}} Q(\mathbf{D}_s - \{X_v\}) \ln \phi_s(\mathbf{D}_s) \right\}$$

where $Z_v = \sum_{X_v} Q(X_v)$ is a normalization constant for this distribution (the sum is replaced with the integral over the value space of X_v if this is a continuous variable), and $\mathbf{D}_s - \{X_v\}$ is the subset of the random variables for the factor s excluding the variable X_v.

Below we specialize the above update formula for each of our variable in our probabilistic model.

1.1 Deformation variables

The mean-field update for each deformation variables is the following:

$$Q(\mathbf{D}_{t,k}) \propto \exp\left\{ -0.5 \sum_p Q(U_{t,p} = k)(\mathbf{D}_{t,k} - \mathbf{X}_{t,p})^T \Sigma_1^{-1}(\mathbf{D}_{t,k} - \mathbf{X}_{t,p}) \right.$$

$$\left. -0.5 \sum_{k' \in N(k)} (\mathbf{D}_{t,k} - \boldsymbol{\mu}_{t,k,k'})^T \Sigma_2^{-1}(\mathbf{D}_{t,k} - \boldsymbol{\mu}_{t,k,k'}) \right\}$$

where $\mathcal{N}(k)$ includes all neighboring points k' of point k on the part template (see main text of the paper) and $\boldsymbol{\mu}_{t,k}$ is a 3D vector defined as follows:

$$\boldsymbol{\mu}_{t,k,k'} = \mathbb{E}_Q[\mathbf{D}_{t,k'}] + (\mathbb{E}_Q[\mathbf{Y}_k] - \mathbb{E}_Q[\mathbf{Y}_{k'}]))$$

We note that the above distribution is a product of Gaussians; when re-normalized, the distribution is equivalent to a Gaussian with the following expectation, or mean, which we use in other updates:

$$\mathbb{E}_Q[\mathbf{D}_{t,k}] = (\sum_p Q(U_{t,p} = k)\Sigma_1^{-1} + \sum_{k' \in N(k)} \Sigma_2^{-1})^{-1}$$

$$\cdot (\sum_p Q(U_{t,p} = k)\Sigma_1^{-1}\mathbf{X}_{t,p} + \sum_{k' \in N(k)} \Sigma_2^{-1}\boldsymbol{\mu}_{t,k,k'}) \tag{1}$$

The above formula indicates that the most likely deformed position of a point on a part template is a weighted average of its associated points on the input surface as well as its neighbors. The weights are controlled by the covariance matrices Σ_1 and Σ_2 as well as the degree of association between the part template point and each input surface point, given by $Q(U_{t,p} = k)$. The covariance matrix of the above distribution is forced to be diagonal (see next section); its diagonal elements tend to increase when the input surface points have weak associations with the part template point, as indicated by the following formula:

$$Cov_Q[\mathbf{D}_{t,k}] = (\sum_p Q(U_{t,p} = k)\Sigma_1^{-1} + \sum_{k' \in N(k)} \Sigma_2^{-1})^{-1}$$

Computing the above expectation and covariance for each variable $\mathbf{D}_{t,k}$ involving summing over every surface point p on the input shape t. This is computationally too expensive. Practically in our implementation, we find the 100 nearest input surface points for each part template point k, and we also find the 20 nearest part template points for each input surface point p. For each template point k, we always keep indices to its 100 nearest surface points as well as the surface points

whose nearest points include that template point k. Instead of summing over all the surface points of each input shape, for each template point k we sum over its abovementioned indices to surface point only. For the rest of the surface points, the distribution values $Q(U_{t,p} = k)$ are practically negligible and are skipped in the above summations.

1.2   Part template variables

For each part template point $\mathbf{Y}_k$, the mean-field update is given by:

$$Q(\mathbf{Y}_k) \propto \exp \left\{ - .5(\mathbf{Y}_k - \boldsymbol{\mu}_k)^T \Sigma_2^{-1}(\mathbf{Y}_k - \boldsymbol{\mu}_k) \right\}$$

where $\boldsymbol{\mu}_k$ is the mean, or expectation (3D vector), defined as follows:

$$\boldsymbol{\mu}_k = \mathbb{E}_Q[\mathbf{Y}_k] = \frac{1}{|\mathcal{N}(k)|} \sum_{k'} \left(\mathbb{E}_Q[\mathbf{Y}_{k'}] + \frac{1}{T} \sum_t (\mathbb{E}_Q[\mathbf{D}_{t,k}] - \mathbb{E}_Q[\mathbf{D}_{t,k'}]) \right)$$

and $\mathcal{N}(k)$ includes all neighboring points k' of point k on the part template, T is the number of input shapes. The covariance matrix for the above distribution is given by Σ_2.

1.3   Point correspondence variables

The mean-field update for the latent variables $U_{t,p}$ yields a categorical distribution computed as follows:

$$\begin{aligned} Q(U_{t,p} = k) \propto \exp \Big\{ &- 0.5(\mathbb{E}_Q[\mathbf{D}_{t,k}] - \mathbf{X}_{t,p})^T \Sigma_1^{-1}(\mathbb{E}_Q[\mathbf{D}_{t,k}] - \mathbf{X}_{t,p}) \\ &- 0.5 Tr(\Sigma_1^{-1} \cdot Cov_Q[\mathbf{D}_{t,k}]) \\ &- 0.5(\mathbf{f}_k - \mathbf{f}_{t,p})^T \Sigma_3^{-1}(\mathbf{f}_k - \mathbf{f}_{t,p}) - \ln \epsilon \cdot Q(S_{t,p} = label(k)) \Big\} \end{aligned}$$

where $Tr(\Sigma_1^{-1} \cdot Cov_Q[\mathbf{D}_{t,k}])$ represents the matrix trace, ϵ is a small constant discussed in the main text of the paper. For computational efficiency reasons, we avoid computing the above distribution for all pairs of part template and input surface points. As in the case of the updates for the deformation variables, we keep indices to input surface point positions that are nearest neighbors to part template points and vice versa. We compute the above distributions only for pairs between these neighboring points. For the rest of the pairs, we set $Q(U_{t,p} = k) = 0$.

1.4 Segmentation variables

The mean-field update for the variables $S_{t,p}$ also yields a categorical distribution:

$$Q(S_{t,p} = l) \propto \exp\Big\{ \sum_k Q(U_{t,p} = k)[label(k) \neq l] \ln \epsilon$$
$$+ \sum_{p' \in N(p)} \sum_{l' \neq l} Q(S_{t,p'} = l') \ln(1.0 - \Phi)$$
$$+ \sum_{p' \in N(p)} Q(S_{t,p'} = l) \ln(\Phi) \Big\}$$

where $N(p)$ is the neighborhood of the input surface point used for segmentation (see main text for more details), Phi evaluates feature differences between neighboring surface points (see main text for its definition). The binary indicator function $[label(k) \neq l]$ is 1 if the expression in brackets holds, otherwise it is 0.

2 Covariance matrix updates

The covariance matrices of our factors are updated as follows:

$$\Sigma_1 = \frac{1}{Z_1} \sum_{t,k,p \in \mathcal{N}(k)} Q(U_{t,p} = k)(\mathbb{E}_Q[\mathbf{D}_{t,k}] - \mathbf{X}_{t,p})(\mathbb{E}_Q[\mathbf{D}_{t,k}] - \mathbf{X}_{t,p})^T$$

$$\Sigma_2 = \frac{1}{Z_2} \sum_{t,k,k' \in N(k)} ((\mathbb{E}_Q[\mathbf{D}_{t,k}] - \mathbb{E}_Q[\mathbf{D}_{t,k'}]) - (\mathbb{E}_Q[\mathbf{Y}_k] - \mathbb{E}_Q[\mathbf{Y}_{k'}])) \cdot$$
$$\cdot (\mathbb{E}_Q[\mathbf{D}_{t,k}] - \mathbb{E}_Q[\mathbf{D}_{t,k'}] - (\mathbb{E}_Q[\mathbf{Y}_k] - \mathbb{E}_Q[\mathbf{Y}_{k'}]))^T$$

$$\Sigma_3 = \frac{1}{Z_3} \sum_{t,k,p \in \mathcal{N}(k)} Q(U_{t,p} = k)(\mathbf{f}_k - \mathbf{f}_{t,p})(\mathbf{f}_k - \mathbf{f}_{t,p})^T$$

$$\Sigma_5 = \frac{1}{Z_5} \sum_{t,p,p' \in N(p)} (\mathbf{f}_{t,p} - \mathbf{f}_{t,p'})(\mathbf{f}_{t,p} - \mathbf{f}_{t,p'})^T$$

where $Z_1 = Z_3 = \sum\limits_{t,k,p \in \mathcal{N}(k)} Q(U_{t,p} = k)$, $Z_2 = \sum\limits_{t,k,k' \in N(k)} 1$, $Z_5 = \sum\limits_{t,p,p' \in N(p)} 1$.
The computed covariance matrices are forced to be diagonal i.e., in the above computations only the diagonal elements of the covariance matrices are taken into account, while the rest of the elements are set to 0.

3 Contrastive divergence

Contrastive divergence iterates over the following three steps in our implementation: variational (mean-field) inference, stochastic approximation (or sampling), and parameter updates. We discuss the steps in detail below.

Variational inference step. Our deformation model yields expectations over deformed point positions of the part templates based on Equation 1. For each deformed point, we find the surface point that is closest to its expected position. Let $D_{k,\tau}[t]$ the observed surface position of point k for an input shape t. The subscript τ takes values 1, 2, or 3 that correspond to the $x-$,$y-$,$z-$ coordinate of the point respectively. Let $E_k[t]$ represents the observed existence of a point k (binary variable) also inferred by our deformation model. Given all observed point positions $\mathbf{D}[t]$ and existences $\mathbf{E}[t]$ per shape t, we perform bottom-up mean-field inference on the binary hidden nodes according to the following equations in the following order:

$$Q(H_m^{(1)} = 1|\mathbf{D}[t], \mathbf{E}[t]) = \sigma \Big(w_{m,0} + \sum_{k \in \mathcal{N}_{m,\tau}} (a_{k,\tau,m} - c_{k,\tau,m}) \ln(D_{k,\tau}[t]) E_k[t]$$
$$+ \sum_{k \in \mathcal{N}_{m,\tau}} (b_{k,\tau,m} - d_{k,\tau,m}) \ln(1 - D_{k,\tau}[t]) E_k[t]$$
$$+ \sum_n w_{m,n} Q(H_n^{(2)} = 1|\mathbf{D}[t], \mathbf{E}[t]) \Big)$$

$$Q(H_n^{(2)} = 1|\mathbf{D}[t], \mathbf{E}[t]) = \sigma \Big(w_{n,0} + \sum_m w_{m,n} Q(H_m^{(1)} = 1|\mathbf{D}[t], \mathbf{E}[t])$$
$$+ \sum_o w_{n,o} Q(H_o^{(3)} = 1|\mathbf{D}[t], \mathbf{E}[t]) \Big)$$

$$Q(H_o^{(3)} = 1|\mathbf{D}[t], \mathbf{E}[t]) = \sigma \Big(w_{o,0} + \sum_n w_{n,o} Q(H_n^{(2)} = 1|\mathbf{D}[t], \mathbf{E}[t]) \Big)$$

where $\sigma(\cdot)$ represents the sigmoid function, $\mathcal{N}_m$ is the set of the observed variables each hidden node (variable) $H_m^{(1)}$ is connected to. The mean-field updates for $H_m^{(1)}$

involve a weighted summation over the observed variables $\mathbf{D}[t]$ per part, which can be thought of as a convolutional scheme per part. We perform 3 mean-field iterations alternating the updates over the above hidden nodes. During the first iteration, we initialize $Q(H_n^{(2)} = 1) = 0$ and $Q(H_o^{(3)} = 1) = 0$ for each hidden node n and o.

Stochastic approximation. This step begins by sampling the binary hidden nodes of the top layer for each training shape t. Sampling is performed according to the inferred distributions $Q(H_o^{(3)} = 1|\mathbf{D}[t], \mathbf{E}[t])$ of the previous step. Let $H_o^{(3)}[t']$ the resulting sampled $0/1$ values. Then we perform top-down mean-field inference on the binary hidden nodes of the other layers as well as the visible layer:

$$Q(H_n^{(2)} = 1|\mathbf{E}[t], \mathbf{H}^{(3)}[t']) = \sigma\bigg(w_{n,0} + \sum_m w_{m,n}Q(H_m^{(1)} = 1|\mathbf{E}[t], \mathbf{H}^{(3)}[t'])$$
$$+ \sum_o w_{n,o}H_o^{(3)}[t']\bigg)$$

$$Q(H_m^{(1)} = 1|\mathbf{E}[t], \mathbf{H}^{(3)}[t']) = \sigma\bigg(w_{m,0} + \sum_{k \in \mathcal{N}_{m,\tau}} (a_{k,\tau,m} - c_{k,\tau,m})\ln(D_{k,\tau}[t'])E_k[t]$$
$$+ \sum_{k \in \mathcal{N}_{m,\tau}} (b_{k,\tau,m} - d_{k,\tau,m})\ln(1 - D_{k,\tau}[t'])E_k[t]$$
$$+ \sum_n w_{m,n}Q(H_n^{(2)} = 1|\mathbf{E}[t], \mathbf{H}^{(3)}[t'])\bigg)$$

$$Q(D_{k,\tau}|\mathbf{E}[t], \mathbf{H}^{(3)}[t']) \propto$$
$$D_{k,\tau}^{(a_{k,\tau,0}-1)+\sum_{m \in \mathcal{N}_k} a_{k,\tau,m}Q(H_m^{(1)}=1|\mathbf{E}[t],\mathbf{H}^{(3)}[t'])+\sum_{m \in \mathcal{N}_k} c_{k,\tau,m}(1-Q(H_m^{(1)}=1|\mathbf{E}[t],\mathbf{H}^{(3)}[t']))}$$
$$(1-D_{k,\tau})^{(b_{k,\tau,0}-1)+\sum_{m \in \mathcal{N}_k} b_{k,\tau,m}Q(H_m^{(1)}=1|\mathbf{E}[t],\mathbf{H}^{(3)}[t'])+\sum_{m \in \mathcal{N}_k} d_{k,\tau,m}(1-Q(H_m^{(1)}=1|\mathbf{E}[t],\mathbf{H}^{(3)}[t']))}$$

where $D_{k,\tau}[t']$ in the above mean-field updates is set to be the expectation of the above Beta distribution and $\mathcal{N}_k$ is the set of the hidden variables each observed node (variable) $D_{k,\tau}$ is connected to. We note that sampling all the variables in the model caused contrastive divergence not to converge, thus we instead used expectations of the above distributions instead. As in the previous step, we performed 3 iterations alternating over the above mean-field updates. During the first iteration, we skip the terms involving $D_{k,\tau}[t']$ during the inference of the hidden nodes of the first layer. At the second and third iteration, we infer distributions for the hidden layers as follows:

$$Q(H_o^{(3)} = 1|\mathbf{D}[t'], \mathbf{E}[t]) = \sigma\bigg(w_{o,0} + \sum_n w_{n,o}Q(H_n^{(2)} = 1|\mathbf{D}[t'], \mathbf{E}[t])\bigg)$$

20

$$Q(H_n^{(2)} = 1|\mathbf{D}[t'], \mathbf{E}[t]) = \sigma\left(w_{n,0} + \sum_m w_{m,n} Q(H_m^{(1)} = 1|\mathbf{D}[t'], \mathbf{E}[t])\right.$$

$$\left. + \sum_o w_{n,o} Q(H_o^{(3)} = 1|\mathbf{D}[t'], \mathbf{E}[t])\right)$$

$$Q(H_m^{(1)} = 1|\mathbf{D}[t'], \mathbf{E}[t]]) = \sigma\left(w_{m,0} + \sum_{k \in \mathcal{N}_{m,\tau}} (a_{k,\tau,m} - c_{k,\tau,m}) \ln(D_{k,\tau}[t']) E_k[t]\right.$$

$$+ \sum_{k \in \mathcal{N}_{m,\tau}} (b_{k,\tau,m} - d_{k,\tau,m}) \ln(1 - D_{k,\tau}[t']) E_k[t]$$

$$\left. + \sum_n w_{m,n} Q(H_n^{(2)} = 1|\mathbf{D}[t'], \mathbf{E}[t])\right)$$

Parameter updates. The parameters of the model are updated with project gradient ascent according to the expectations over the final distributions computed in the previous two steps and the observed data. We list the parameter updates below. We note that $sgn(\cdot)$ used below denotes the sign function, ν is the iteration number (or epoch), η is the learning rate, μ is the momentum rate. The learning rate is set to 0.001 initially, and is multiplied by a factor 0.9 when at the previous epoch the reconstruction error $\sum_{t,k,\tau} |D_{k,\tau}[t]E_k[t] - D_{k,\tau}[t']E_k[t]|$ increases, and it is multiplied by a factor 1.1 when the reconstruction error decreases. The momentum rate is progressively increased from 0.5 towards 1.0 asymptotically during training.

$$a_{k,\tau,0} = max(a_{k,\tau,0} + \Delta a_{k,\tau,0}[\nu], 0) \quad \text{, where}$$

$$\Delta a_{k,\tau,0}[\nu] = \mu \cdot \Delta a_{k,\tau,0}[\nu - 1] + \eta \frac{1}{T} \sum_t \left(\ln(D_{k,\tau}[t]) - \ln(D_{k,\tau}[t'])\right)$$

$$- \eta\lambda_1 \sum_{k' \in \mathcal{N}_k} sgn(a_{k,\tau,0} - a_{k',\tau,0}) - \eta\lambda_2 sgn(a_{k,\tau,0})$$

$$b_{k,\tau,0} = max(b_{k,\tau,0} + \Delta b_{k,\tau,0}[\nu], 0) \quad \text{, where}$$

$$\Delta b_{k,\tau,0}[\nu] = \mu \cdot \Delta b_{k,\tau,0}[\nu - 1] + \eta \frac{1}{T} \sum_t \left(\ln(1 - D_{k,\tau}[t]) - \ln(1 - D_{k,\tau}[t'])\right)$$

$$- \eta\lambda_1 \sum_{k' \in \mathcal{N}_k} sgn(b_{k,\tau,0} - b_{k',\tau,0}) - \eta\lambda_2 sgn(b_{k,\tau,0})$$

$$a_{k,\tau,m} = max(a_{k,\tau,m} + \Delta a_{k,\tau,m}[\nu], 0) \quad \text{, where}$$

$$\Delta a_{k,\tau,m}[\nu] = \mu \cdot \Delta a_{k,\tau,m}[\nu-1] + \eta \frac{1}{T} \sum_t \left(\ln(D_{k,\tau}[t])Q(H_m^{(1)} = 1|\mathbf{D}[t], \mathbf{E}[t]) \right.$$

$$\left. - \ln(D_{k,\tau}[t'])Q(H_m^{(1)} = 1|\mathbf{D}[t'], \mathbf{E}[t]) \right)$$

$$- \eta \lambda_1 \sum_{k' \in \mathcal{N}_k} sgn(a_{k,\tau,m} - a_{k',\tau,m}) - \eta \lambda_2 sgn(a_{k,\tau,m})$$

$$b_{k,\tau,m} = max(b_{k,\tau,m} + \Delta b_{k,\tau,m}[\nu], 0) \quad \text{, where}$$

$$\Delta b_{k,\tau,m}[\nu] = \mu \cdot \Delta b_{k,\tau,m}[\nu-1] + \eta \frac{1}{T} \sum_t \left(\ln(1 - D_{k,\tau}[t])Q(H_m^{(1)} = 1|\mathbf{D}[t], \mathbf{E}[t]) \right.$$

$$\left. - \ln(1 - D_{k,\tau}[t'])Q(H_m^{(1)} = 1|\mathbf{D}[t'], \mathbf{E}[t]) \right)$$

$$- \eta \lambda_1 \sum_{k' \in \mathcal{N}_k} sgn(b_{k,\tau,m} - b_{k',\tau,m}) - \eta \lambda_2 sgn(b_{k,\tau,m})$$

$$c_{k,\tau,m} = max(c_{k,\tau,m} + \Delta c_{k,\tau,m}[\nu], 0) \quad \text{, where}$$

$$\Delta c_{k,\tau,m}[\nu] = \mu \cdot \Delta c_{k,\tau,m}[\nu-1] + \eta \frac{1}{T} \sum_t \left(\ln(D_{k,\tau}[t])(1 - Q(H_m^{(1)} = 1|\mathbf{D}[t], \mathbf{E}[t]) \right.$$

$$\left. - \ln(D_{k,\tau}[t'])(1 - Q(H_m^{(1)} = 1|\mathbf{D}[t'], \mathbf{E}[t]) \right)$$

$$- \eta \lambda_1 \sum_{k' \in \mathcal{N}_k} sgn(c_{k,\tau,m} - c_{k',\tau,m}) - \eta \lambda_2 sgn(c_{k,\tau,m})$$

$$d_{k,\tau,m} = max(d_{k,\tau,m} + \Delta d_{k,\tau,m}[\nu], 0) \quad \text{, where}$$

$$\Delta d_{k,\tau,m}[\nu] = \mu \cdot \Delta d_{k,\tau,m}[\nu - 1] + \eta \frac{1}{T} \sum_t \left(\ln(1 - D_{k,\tau}[t])(1 - Q(H_m^{(1)} = 1|\mathbf{D}[t], \mathbf{E}[t]) \right.$$

$$\left. - \ln(1 - D_{k,\tau}[t'])(1 - Q(H_m^{(1)} = 1|\mathbf{D}[t'], \mathbf{E}[t]) \right)$$

$$- \eta \lambda_1 \sum_{k' \in \mathcal{N}_k} sgn(d_{k,\tau,m} - d_{k',\tau,m}) - \eta \lambda_2 sgn(d_{k,\tau,m})$$

$$w_{m,0} = w_{m,0} + \Delta w_{m,0}[\nu] \quad, \text{where}$$

$$\Delta w_{m,0}[\nu] = \mu \cdot \Delta w_{m,0}[\nu - 1] + \eta \frac{1}{T} \sum_t \left(Q(H_m^{(1)} = 1|\mathbf{D}[t], \mathbf{E}[t]) \right.$$

$$\left. - Q(H_m^{(1)} = 1|\mathbf{D}[t'], \mathbf{E}[t]) \right)$$

$$- \eta \lambda_2 sgn(w_{m,0})$$

$$w_{n,0} = w_{n,0} + \Delta w_{n,0}[\nu] \quad, \text{where}$$

$$\Delta w_{n,0}[\nu] = \mu \cdot \Delta w_{n,0}[\nu - 1] + \eta \frac{1}{T} \sum_t \left(Q(H_n^{(2)} = 1|\mathbf{D}[t], \mathbf{E}[t]) \right.$$

$$\left. - Q(H_n^{(2)} = 1|\mathbf{D}[t'], \mathbf{E}[t]) \right)$$

$$- \eta \lambda_2 sgn(w_{n,0})$$

$$w_{o,0} = w_{o,0} + \Delta w_{o,0}[\nu] \quad, \text{where}$$

$$\Delta w_{o,0}[\nu] = \mu \cdot \Delta w_{o,0}[\nu - 1] + \eta \frac{1}{T} \sum_t \left(Q(H_o^{(3)} = 1|\mathbf{D}[t], \mathbf{E}[t]) \right.$$

$$\left. - Q(H_o^{(3)} = 1|\mathbf{D}[t'], \mathbf{E}[t]) \right)$$

$$- \eta \lambda_2 sgn(w_{o,0})$$

$$w_{m,n} = w_{m,n} + \Delta w_{m,n}[\nu] \quad , \text{ where}$$

$$\Delta w_{m,n}[\nu] = \mu \cdot \Delta w_{m,n}[\nu - 1] + \eta \frac{1}{T} \sum_t \left(Q(H_m^{(1)} = 1|\mathbf{D}[t], \mathbf{E}[t]) Q(H_n^{(2)} = 1|\mathbf{D}[t], \mathbf{E}[t]) \right.$$
$$\left. - Q(H_m^{(1)} = 1|\mathbf{D}[t'], \mathbf{E}[t]) Q(H_n^{(2)} = 1|\mathbf{D}[t'], \mathbf{E}[t]) \right)$$
$$- \eta \lambda_2 sgn(w_{m,n})$$

$$w_{n,o} = w_{n,o} + \Delta w_{n,o}[\nu] \quad , \text{ where}$$

$$\Delta w_{n,o}[\nu] = \mu \cdot \Delta w_{n,o}[\nu - 1] + \eta \frac{1}{T} \sum_t \left(Q(H_n^{(2)} = 1|\mathbf{D}[t], \mathbf{E}[t]) Q(H_o^{(3)} = 1|\mathbf{D}[t], \mathbf{E}[t]) \right.$$
$$\left. - Q(H_n^{(2)} = 1|\mathbf{D}[t'], \mathbf{E}[t]) Q(H_o^{(3)} = 1|\mathbf{D}[t'], \mathbf{E}[t]) \right)$$
$$- \eta \lambda_2 sgn(w_{n,o})$$

3.1 Parameter updates for the structure part of the BSM

To learn the parameters $w_{k,0}$, $w_{k,r}$, $w_{r,0}$ involving the variables $\mathbf{E}$ and $\mathbf{G}$ of the BSM part modeling the shape structure, we similarly perform contrastive divergence with the following steps:

Inference. Given the observed point existences $\mathbf{E}[t]$, we infer the following distribution over the latent variables $\mathbf{G}$ (we note that this is exact inference):

$$Q(G_r = 1|\mathbf{E}[t]) = \sigma(w_{r,0} + \sum_k w_{k,r} E_k[t])$$

Sampling. We sample the binary latent variables $\mathbf{G}$ according to the inferred distribution $Q(G_r = 1|\mathbf{E}[t])$. Let $G_r[t']$ the resulting sampled 0/1 values. We perform inference for the existence variables as follows:

$$Q(E_k = 1|\mathbf{G}[t']) = \sigma(w_{k,0} + \sum_r w_{k,r} G_r[t'])$$

24

and repeat for the latent variables:

$$Q(G_r = 1|\mathbf{G}[t']) = \sigma(w_{r,0} + \sum_k w_{k,r}Q(E_k = 1|\mathbf{G}[t']))$$

Parameter updates The parameters of the structure part of the BSM are updated as follows:

$$w_{k,0} = w_{k,0} + \Delta w_{k,0}[\nu] \quad \text{, where}$$

$$\Delta w_{k,0}[\nu] = \mu \cdot \Delta w_{k,0}[\nu - 1] + \eta \frac{1}{T} \sum_t \left(E_k[t] - Q(E_k = 1|\mathbf{G}[t']) \right)$$
$$- \eta\lambda_1 \sum_{k' \in \mathcal{N}_k} sgn(w_{k,0} - w_{k',0}) - \eta\lambda_2 sgn(w_{k,0})$$

$$w_{r,0} = w_{r,0} + \Delta w_{r,0}[\nu] \quad \text{, where}$$

$$\Delta w_{r,0}[\nu] = \mu \cdot \Delta w_{r,0}[\nu - 1] + \eta \frac{1}{T} \sum_t \left(Q(G_r = 1|\mathbf{E}[t]) - Q(G_r = 1|\mathbf{G}[t']) \right)$$
$$- \eta\lambda_2 sgn(w_{r,0})$$

$$w_{k,r} = w_{k,r} + \Delta w_{k,r}[\nu] \quad \text{, where}$$

$$\Delta w_{k,r}[\nu] = \mu \cdot \Delta w_{k,r}[\nu - 1] + \eta \frac{1}{T} \sum_t \left(E_k[t]Q(G_r = 1|\mathbf{E}[t]) \right.$$
$$\left. - Q(E_k = 1|\mathbf{G}[t'])Q(G_r = 1|\mathbf{G}[t']) \right)$$
$$- \eta\lambda_1 \sum_{k' \in \mathcal{N}_k} sgn(w_{k,r} - w_{k',r}) - \eta\lambda_2 sgn(w_{k,r})$$

References

[1] Daphne Koller and Nir Friedman. *Probabilistic Graphical Models: Principles and Techniques.* The MIT Press, 2009.