Supplemental Material
A Mixture of Manhattan Frames: Beyond the Manhattan World

Julian Straub Guy Rosman Oren Freifeld John J. Leonard John W. Fisher III
Massachusetts Institute of Technology
{jstraub,rosman,freifeld,jleonard,fisher}@csail.mit.edu

Abstract

This is the supplemental material for the paper "A Mixture of Manhattan Frames: Beyond the Manhattan World" [8]. We derive the distributions involved in the split/merge proposals for the Mixture of Manhattan Frames (MMF). Additionally, we give details for how we define a Gaussian distribution on rotations in 3D, as well as details for the Riemannian exponential and logarithm operators for the manifold of the unit sphere S^2.

1. Split/Merge Proposals for the Mixture of Manhattan Frames (MMF) Model

Here we derive split and merge proposals for the MMF model as well as their acceptance probability in an approach similar to Richardson and Green [7]. Note that a merge involves moving all points from MF l and m into a new MF n and then removing MFs l and m. Similarly, a split creates two new MFs l and m from a single MF n. Hence, both a split and a merge change the number of parameters in the model. Specifically the parameters that change their dimension are the set of MF rotations, $\mathbf{R}$, and the set of covariances on the MF axes, $\mathbf{\Sigma}$. The labels $\mathbf{z}$ and $\mathbf{c}$ remain the of same dimensions, only the range for $\mathbf{c}$ changes from $[1, K]$ to $[1, K-1]$ (merge) or from $[1, K]$ to $[1, K+1]$ (split). Therefore, we employ the theory of Reverse Jump Markov Chain Monte Carlo (RJMCMC) [4] to derive a proper acceptance probability. RJMCMC is a generalization of Metropolis-Hastings MCMC [5] and provides a way of computing an acceptance probability when the number of parameters changes between moves. In the following we first introduce RJMCMC before we derive split/merge proposal distributions for the MMF as well as the respective acceptance probabilities.

We will see that the split/merge proposals as well as the acceptance probabilities are similar to what one would expect to see when employing the Metropolis-Hastings algorithm. For this reason and because the MH algorithm is more well-known, we chose to refer to the inference algorithm in the MMF paper [8] as to Metropolis-Hastings MCMC.

1.1. General Introduction of Reverse Jump MCMC

RJMCMC utilizes auxiliary variables to propose deterministic moves to change between model orders. In general the RJMCMC algorithm executes the following steps when proposing a move:

1. Draw auxiliary variables $\mathbf{v}$ given the current state $\mathbf{x}$ from some proposal distribution $q(\mathbf{v}|\mathbf{x})$:
$$\mathbf{v} \sim q(\mathbf{v}|\mathbf{x})\,. \tag{1}$$
2. Apply a deterministic function $f([\mathbf{x}, \mathbf{v}])$ to generate the state after the move $\hat{\mathbf{x}}$ as well as auxiliary variables $\mathbf{u}$:
$$f([\mathbf{x}, \mathbf{v}]) = [\mathbf{u}, \hat{\mathbf{x}}]\,. \tag{2}$$
By $[\mathbf{x}, \mathbf{v}]$ we denote stacking of all parameters in $\mathbf{x}$ and $\mathbf{v}$.
3. Accept the move from state $\mathbf{x}$ to state $\hat{\mathbf{x}}$ with probability
$$\Pr\left(\text{accept move}\right) = \min\left\{1, \frac{p(\hat{\mathbf{x}})}{p(\mathbf{x})}\frac{q(\mathbf{x}|\hat{\mathbf{x}})}{q(\hat{\mathbf{x}}|\mathbf{x})}|\det(J_f)|\right\}\,, \tag{3}$$
where the Jacobian $J_f = \frac{\partial f([\mathbf{x},\mathbf{v}])}{\partial[\mathbf{x},\mathbf{v}]}$.

As an example, suppose that we wanted to merge two clusters A and B parameterized by $\mathbf{x}_A$ and $\mathbf{x}_B$ into one clusters C (with parameters $\hat{\mathbf{x}}_C$). We would first sample an auxiliary cluster $\mathbf{v}$ from the current clusters:

$$\mathbf{v} \sim q(\mathbf{v}|\mathbf{x}_A, \mathbf{x}_B)\,, \tag{4}$$

where $q(\mathbf{v}|\mathbf{x}_A, \mathbf{x}_B)$ is some proposal distribution. Second, we would use a deterministic function $f([\mathbf{x}_A, \mathbf{x}_B, \mathbf{v}])$ to generate the parameters for the merged cluster $\hat{\mathbf{x}}_C$. In this example, the deterministic function assigns the auxiliary cluster $\mathbf{v}$ to the merged cluster C:

$$[\mathbf{u}_1, \mathbf{u}_2, \hat{\mathbf{x}}_C] = f([\mathbf{x}_A, \mathbf{x}_B, \mathbf{v}]) = [\mathbf{x}_A, \mathbf{x}_B, \mathbf{v}]\,. \tag{5}$$

Therefore, the Jacobian J_f is the identity matrix which has determinant 1. Merge moves are thus accepted with probability

$$\Pr\left(\text{accept merge}\right) = \min\left\{1, \frac{p(\hat{\mathbf{x}}_C)}{p(\mathbf{x}_A, \mathbf{x}_B)} \frac{q(\mathbf{x}_A, \mathbf{x}_B|\hat{\mathbf{x}}_C)}{q(\hat{\mathbf{x}}_C|\mathbf{x}_A, \mathbf{x}_B)}\right\}\,. \tag{6}$$

1.2. RJMCMC Merge Moves in an MMF

In the following, we will give the RJMCMC algorithm for a merge proposal between two MFs as well as the acceptance probability of the move. Note that the computation of the acceptance probability requires the probability of an inverse proposal: the split proposal $q(\text{split})$. Since the details of the split proposal are not important for the derivation in this section, we defer those details to the next section.

As outlined in the previous section, an RJMCMC merge proposal is executed in the following three steps:

1. Draw an auxiliary MF v parameterized by $\mathbf{v} = \{\mathbf{c}^v, \mathbf{z}^v, w^v, R_v, \mathbf{\Sigma}_{v,\{1\ldots6\}}\}$ from the current MFs l and m parameterized by $\mathbf{x}_l = \{\mathbf{c}_{\{\mathbf{c}=l\}}, \mathbf{z}_{\{\mathbf{c}=l\}}, w_l, R_l, \mathbf{\Sigma}_{l,\{1\ldots6\}}\}$ and $\mathbf{x}_m = \{\mathbf{c}_{\{\mathbf{c}=m\}}, \mathbf{z}_{\{\mathbf{c}=m\}}, w_m, R_m, \mathbf{\Sigma}_{m,\{1\ldots6\}}\}$, where $\mathbf{z}_{\{\mathbf{c}=m\}}$ denotes the set of labels z_i for which $c_i = m$.

   We first assign all normals of MF l and m to MF v: $\mathbf{c}^v_{\mathbf{c}\in\{l,m\}} = v$, which corresponds to the proposal distribution:

   $$q(\mathbf{c}^v_{\{\mathbf{c}\in\{l,m\}\}}|\mathbf{c}) = \delta(\mathbf{c}^v_{\{\mathbf{c}\in\{l,m\}\}} - v)\,. \tag{7}$$

   Second, we sample the axes assignments $\mathbf{z}^v_{\{\hat{\mathbf{c}}=v\}} = \{z^v_i\}_{i:\hat{c}_i=v}$ according to

   $$q(z^v_i = j|w_l, R_l, \mathbf{\Sigma}_l, \mathbf{q}) \quad\propto\quad w_{lj}\, p(q_i; [M_l]_j, \Sigma_{lj})\,. \tag{8}$$

   Next, given associations $\mathbf{c}^v$ and $\mathbf{z}^v$, we optimize the rotation of MF v to obtain $R^\star_v$ starting (arbitrarily) from rotation R_l – in our experiments the conjugate gradient optimization was not sensitive to initialization and R_l as well as R_n were generally close to $R^\star_v$. Then we sample R_v from a narrow Gaussian distribution over rotations with mean $R^\star_v$:

   $$q(R_v|\mathbf{z}^v, \mathbf{c}^v, \mathbf{q}, R_l) = \mathcal{N}(R_v; R^\star_v(\mathbf{z}^v, \mathbf{c}^v, \mathbf{q}, R_l), \Sigma_{\mathrm{so}(3)}) = \mathcal{N}((R^{\star T}_v \mathrm{Log}_{R^\star_v(\cdot)}(R_v))^\vee; 0, \Sigma_{\mathrm{so}(3)})\,, \tag{9}$$

   where $\mathrm{Log}_{R^\star_v}(R) : \mathrm{SO}(3) \to T_{R^\star_v}\mathrm{SO}(3)$ denotes the logarithm map of R into the tangent space $T_{R^\star_v}\mathrm{SO}(3)$ around $R^\star_v$. The vee operator $^\vee$ [2] extracts the unique elements of a skew-symmetric matrix $W \in \mathbb{R}^{3\times3}$ into a vector w: $W^\vee = w = [-W_{23}; W_{13}; -W_{12}] \in \mathbb{R}^3$. $\Sigma_{\mathrm{so}(3)} \in \mathbb{R}^{3\times3}$ is the covariance of the Normal distribution in $T_{R_\mu}\mathrm{SO}(3)$. Refer to Sec. 2 for an in depth discussion.

   Finally, we obtain samples for the axis covariances $\mathbf{\Sigma}_{v,\{1\ldots6\}}$ according to the proposal distribution

   $$q(\mathbf{\Sigma}_{v,\{1\ldots6\}}|\mathbf{c}^v, \mathbf{z}^v, R_v, \mathbf{q}; \Delta, \nu) = \prod_{j=1}^{6} p(\Sigma_{vj}|\mathbf{z}^v, \mathbf{c}^v, \mathbf{q}, R_v; \Delta, \nu)\,, \tag{10}$$

   where $p(\Sigma_{vj}|\mathbf{z}^v, \mathbf{c}^v, \mathbf{q}, R_v; \Delta, \nu)$ is the posterior distribution over covariance Σ_{vj} under the assigned normals in the tangent space $T_{[M_v]_j}S^2$. Since the Inverse Wishart (IW) prior on the covariances is a conjugate prior, the posterior distribution is an IW distribution as well. Therefore, it is straight forward to sample from the posterior distribution.

2. Apply the deterministic function $f([\mathbf{x}_l, \mathbf{x}_m, \mathbf{v}]) = [\mathbf{u}_1, \mathbf{u}_2, \hat{\mathbf{x}}_n]$ to obtain MF n parameterized by $\hat{\mathbf{x}}_n$ (MF after the merge). The auxiliary MFs $\mathbf{u}_1$ and $\mathbf{u}_2$ absorb MFs l and m (MFs before the merge). The function $f([\mathbf{x}_l, \mathbf{x}_m, \mathbf{v}])$ is hence defined as

$$\begin{aligned} \mathbf{u}_1 &= \mathbf{x}_l \\ \mathbf{u}_2 &= \mathbf{x}_m \\ \hat{\mathbf{x}}_n &= \mathbf{v}\,. \end{aligned} \tag{11}$$

Therefore, the Jacobian J_f of the function $f([\mathbf{x}_l, \mathbf{x}_m, \mathbf{v}])$ is

$$J_f = \frac{\partial f([\mathbf{x}_l, \mathbf{x}_m, \mathbf{v}])}{\partial [\mathbf{x}_l, \mathbf{x}_m, \mathbf{v}]} = \mathrm{I}\,, \tag{12}$$

where I is the identity matrix.

3. Accept the merge move with acceptance probability

$$\Pr\left(\text{accept merge}\right) = \min\left\{1, \frac{p(\mathbf{q}, \hat{\mathbf{c}}, \hat{\mathbf{z}}, \hat{\pi}, \hat{\mathbf{w}}, \hat{\mathbf{\Sigma}}, \hat{\mathbf{R}}; \alpha, \gamma, \Delta, \nu)}{p(\mathbf{q}, \mathbf{c}, \mathbf{z}, \pi, \mathbf{w}, \mathbf{\Sigma}, \mathbf{R}; \alpha, \gamma, \Delta, \nu)} \frac{q(\text{split})}{q(\text{merge})} |\det(J_f)|\right\}\,, \tag{13}$$

where parameters after the merge are designated with a hat and J_f is the Jacobian of the deterministic transformation $f(\cdot)$ in the RJMCMC algorithm. The proposal distributions for a split or a merge are denoted $q(\text{split})$ and respectively $q(\text{merge})$. From Eq. (12) we immediately see, that $|\det(J_f)| = 1$ and hence

$$\Pr\left(\text{accept merge}\right) = \min\left\{1, \frac{p(\mathbf{q}, \hat{\mathbf{c}}, \hat{\mathbf{z}}, \hat{\pi}, \hat{\mathbf{w}}, \hat{\mathbf{\Sigma}}, \hat{\mathbf{R}}; \alpha, \gamma, \Delta, \nu)}{p(\mathbf{q}, \mathbf{c}, \mathbf{z}, \pi, \mathbf{w}, \mathbf{\Sigma}, \mathbf{R}; \alpha, \gamma, \Delta, \nu)} \frac{q(\text{split})}{q(\text{merge})}\right\}\,. \tag{14}$$

Note that the RJMCMC acceptance probability for split/merge moves in an MMF looks like the Metropolis-Hastings acceptance probability, because $|\det(J_f)| = 1$. However, since the model orders in the nominator and denominator of the fractions are different, it technically is not a Metropolis-Hastings acceptance probability.

1.3. RJMCMC Split Moves in an MMF

For a split move, the reverse of a merge, we sample two auxiliary MFs $\mathbf{u}_1$ and $\mathbf{u}_2$ and use the inverse of Eq. (11) to obtain the MMF after the split. In the following, we compress steps (1) and (2) in the RJMCMC algorithm into a single: we directly draw the state after the split from the MMF state before. Note that while this process looks like Metropolis-Hastings MCMC, it technically still is RJMCMC since we are changing the model order between moves. Here we denote variables after the split with a hat.

First, we randomly assign normals in MF n to MF l or m by drawing MF assignments to MFs m and l according to

$$q(\hat{\mathbf{c}}_{\{\mathbf{c}=n\}}|\mathbf{c}; \alpha) = \text{DirMult}\left(\hat{\mathbf{c}}_{\{\mathbf{c}=n\}}; \alpha_l, \alpha_m\right)\,. \tag{15}$$

Within each of the MFs l and m we assign normals $\mathbf{q}$ to an axis by drawing the assignments $\hat{\mathbf{z}}_{\{\mathbf{c}=n\}} = \{\hat{z}_i\}_{i:c_i=n}$ as

$$q(\hat{z}_i = j|w_n, R_n, \mathbf{\Sigma}_n, \mathbf{q}) \propto w_{nj}\, p(q_i; [M_n]_j, \Sigma_{nj})\,. \tag{16}$$

Using these assignments, we find optimal rotations $\hat{R}_l^\star$ and $\hat{R}_m^\star$ and draw $\hat{R}_l$ and $\hat{R}_m$:

$$q(\hat{R}_l, \hat{R}_m|\hat{\mathbf{z}}, \hat{\mathbf{c}}, \mathbf{q}, R_n) = \mathcal{N}(\hat{R}_l; \hat{R}_l^\star(R_n, \hat{\mathbf{z}}, \hat{\mathbf{c}}, \mathbf{q}), \Sigma_{\text{so}(3)})\,\mathcal{N}(\hat{R}_m; \hat{R}_m^\star(R_n, \hat{\mathbf{z}}, \hat{\mathbf{c}}, \mathbf{q}), \Sigma_{\text{so}(3)})\,. \tag{17}$$

Given rotations as well as labels, we can draw axis covariances $\hat{\mathbf{\Sigma}}_{\{l,m\},\{1\ldots6\}}$ from the respective posterior:

$$q(\hat{\mathbf{\Sigma}}_{\{l,m\},\{1\ldots6\}}|\hat{\mathbf{c}}, \hat{\mathbf{z}}, \mathbf{q}, \hat{R}_{\{l,m\}}; \Delta, \nu) = \prod_{j=1}^{6} p(\hat{\Sigma}_{lj}|\hat{\mathbf{z}}, \hat{\mathbf{c}}, \mathbf{q}, \hat{R}_l; \Delta, \nu) p(\hat{\Sigma}_{mj}|\hat{\mathbf{z}}, \hat{\mathbf{c}}, \mathbf{q}, \hat{R}_m; \Delta, \nu) \tag{18}$$

While for a merge move the acceptance probability is equal to Eq. (14), we invert the ratio to obtain the acceptance probability for a split move:

$$\Pr\left(\text{accept split}\right) = \min\left\{1, \frac{p(\mathbf{q}, \hat{\mathbf{c}}, \hat{\mathbf{z}}, \hat{\pi}, \hat{\mathbf{w}}, \hat{\mathbf{\Sigma}}, \hat{\mathbf{R}}; \alpha, \gamma, \Delta, \nu)}{p(\mathbf{q}, \mathbf{c}, \mathbf{z}, \pi, \mathbf{w}, \mathbf{\Sigma}, \mathbf{R}; \alpha, \gamma, \Delta, \nu)} \frac{q(\text{merge})}{q(\text{split})}\right\}\,. \tag{19}$$

1.4. Acceptance Probability for Split/Merge Proposals in an MMF

After introducing the RJMCMC merge and the split proposals in the previous sections, we will now derive the acceptance probabilities for those two moves by detailing the distributions involved in the computation of Eq. (14) and (19). First, the joint distribution for the MMF model is

$$p(\mathbf{q}, \mathbf{c}, \mathbf{z}, \pi, \mathbf{w}, \mathbf{\Sigma}, \mathbf{R}; \alpha, \gamma, \Delta, \nu) = \begin{array}{l} \prod_i^N p(q_i|c_i, z_i, R_{c_i}, \Sigma_{c_i z_i}) p(z_i|c_i, w_{c_i}) p(c_i|\pi) p(\pi; \alpha) \prod_k^K p(R_k) p(w_k; \gamma) \\ \prod_j^6 p(\Sigma_{kj}; \Delta, \nu)\,. \end{array} \tag{20}$$

For the evaluation of the acceptance probability, we marginalize over the categorical variables π and $\{w_k\}_{k=1}^K$:

$$p(\mathbf{c}; \alpha) = \int_\pi p(\mathbf{c}|\pi) p(\pi; \alpha) \mathrm{d}\pi = \mathrm{DirMult}(\mathbf{c}; \alpha) \tag{21}$$

$$p(\mathbf{z}_{\{\mathbf{c}=k\}}|\mathbf{c}; \gamma) = \int_{w_k} p(\mathbf{z}_{\{\mathbf{c}=k\}}|\mathbf{c}, w_k) p(w_k; \gamma) \mathrm{d}w_k = \mathrm{DirMult}(\mathbf{z}_{\{\mathbf{c}=k\}}; \gamma)\,, \tag{22}$$

where $\mathbf{z}_{\{\mathbf{c}=k\}} = \{z_i\}_{i:c_i=k}$ and DirMult denotes the Dirichlet Multinomial distribution:

$$\mathrm{DirMult}(\mathbf{c}; \alpha) = \frac{\Gamma(\sum_{k=1}^K \alpha_k)}{\Gamma(\sum_{k=1}^K \alpha_k + N_k)} \prod_{k=1}^K \frac{\Gamma(\alpha_k + N_k)}{\Gamma(\alpha_k)}\,. \tag{23}$$

The counts N_k of labels $\mathbf{c}$ pointing to MF k are computed as $N_k = \sum_{i=1}^N \mathbb{1}_{[c_i=k]}$. After marginalization of π and $\{w_k\}_{k=1}^K$, the joint distribution of the MMF is:

$$p(\mathbf{q}, \mathbf{c}, \mathbf{z}, \mathbf{\Sigma}, \mathbf{R}; \alpha, \gamma, \Delta, \nu) = \quad p(\mathbf{c}; \alpha) \prod_i^N p(q_i|c_i, z_i, R_{c_i}, \Sigma_{c_i z_i}) \prod_k^K p(R_k) p(\mathbf{z}_{\{\mathbf{c}=k\}}|\mathbf{c}; \gamma) \prod_j^6 p(\Sigma_{kj}; \Delta, \nu)\,, \tag{24}$$

where we have assumed that the prior over rotations factors according to $p(\mathbf{R}) = \prod_{k=1}^K p(R_k)$. Each $p(R_k)$ is an uniform distribution over all rotations $R_k \in \mathrm{SO}(3)$. Since $\mathrm{SO}(3)$ is a manifold with finite support, we can compute its volume and obtain $8\pi^2$ [2] which implies $p(R_k) = \frac{1}{8\pi^2}$. Therefore, the ratio of joint probabilities in the merge move acceptance probability in Eq. (14) becomes

$$\frac{p(\mathbf{q},\hat{\mathbf{c}},\hat{\mathbf{z}},\hat{\mathbf{\Sigma}},\hat{\mathbf{R}};\alpha,\gamma,\Delta,\nu)}{p(\mathbf{q},\mathbf{c},\mathbf{z},\mathbf{\Sigma},\mathbf{R};\alpha,\gamma,\Delta,\nu)} = \frac{p(\hat{\mathbf{c}};\alpha)p(\mathbf{q}|\hat{\mathbf{c}},\hat{\mathbf{z}},\hat{\mathbf{R}},\hat{\mathbf{\Sigma}})p(\hat{\mathbf{z}}|\hat{\mathbf{c}};\gamma)p(\hat{\mathbf{\Sigma}};\Delta,\nu)p(\hat{\mathbf{R}})}{p(\mathbf{c};\alpha)p(\mathbf{q}|\mathbf{c},\mathbf{z},\mathbf{R},\mathbf{\Sigma})p(\mathbf{z}|\mathbf{c};\gamma)p(\mathbf{\Sigma};\Delta,\nu)p(\mathbf{R})} = \frac{8\pi^2 p(\hat{\mathbf{c}};\alpha)\big(\prod_i^N p(q_i|\hat{c}_i,\hat{z}_i,\hat{\mathbf{R}},\hat{\mathbf{\Sigma}})\big)\prod_{k=1}^{\hat{K}} p(\hat{\mathbf{z}}_{\{\hat{\mathbf{c}}=k\}}|\hat{\mathbf{c}};\gamma)\prod_{j=1}^6 p(\hat{\Sigma}_{kj};\Delta,\nu)}{p(\mathbf{c};\alpha)\big(\prod_i^N p(q_i|c_i,z_i,\mathbf{R},\mathbf{\Sigma})\big)\prod_{k=1}^K p(\mathbf{z}_{\{\mathbf{c}=k\}}|\mathbf{c};\gamma)\prod_{j=1}^6 p(\Sigma_{kj};\Delta,\nu)}\,, \tag{25}$$

where $\hat{K} = K - 1$. For the acceptance probability of a split proposal in Eq. (19) we get the following ratio:

$$\frac{p(\mathbf{q},\hat{\mathbf{c}},\hat{\mathbf{z}},\hat{\mathbf{\Sigma}},\hat{\mathbf{R}};\alpha,\gamma,\Delta,\nu)}{p(\mathbf{q},\mathbf{c},\mathbf{z},\mathbf{\Sigma},\mathbf{R};\alpha,\gamma,\Delta,\nu)} = \frac{p(\hat{\mathbf{c}};\alpha)\big(\prod_i^N p(q_i|\hat{c}_i,\hat{z}_i,\hat{\mathbf{R}},\hat{\mathbf{\Sigma}})\big)\prod_{k=1}^{\hat{K}} p(\hat{\mathbf{z}}_{\{\hat{\mathbf{c}}=k\}}|\hat{\mathbf{c}};\gamma)\prod_{j=1}^6 p(\hat{\Sigma}_{kj};\Delta,\nu)}{8\pi^2 p(\mathbf{c};\alpha)\big(\prod_i^N p(q_i|c_i,z_i,\mathbf{R},\mathbf{\Sigma})\big)\prod_{k=1}^K p(\mathbf{z}_{\{\mathbf{c}=k\}}|\mathbf{c};\gamma)\prod_{j=1}^6 p(\Sigma_{kj};\Delta,\nu)}\,, \tag{26}$$

where $\hat{K} = K + 1$.

For a split or a merge move, we have to sample new labels, new rotations as well as new covariances given the assignments. Therefore, we formulate the split and merge proposals as

$$q(\text{split}) = q(\mathbf{c}, \mathbf{z}, w_{\{l,m\}}, R_{\{l,m\}}, \mathbf{\Sigma}_{\{l,m\},\{1\ldots6\}}|\hat{\mathbf{c}}, \hat{w}_n, \hat{R}_n, \hat{\mathbf{\Sigma}}_n, \mathbf{q}; \alpha, \gamma, \Delta, \nu) \tag{27}$$

$$q(\text{merge}) = q(\hat{\mathbf{c}}, \hat{\mathbf{z}}, \hat{w}_n, \hat{R}_n, \hat{\mathbf{\Sigma}}_{n,\{1\ldots6\}}|\mathbf{c}, w_{\{l,m\}}, R_{\{l,m\}}, \mathbf{\Sigma}_{\{l,m\}}, \mathbf{q}; \alpha, \gamma, \Delta, \nu)\,. \tag{28}$$

In the following we will first derive the merge and then the split proposal distributions.

The merge proposal is the same as derived previously for the auxiliary MF $\mathbf{v}$ in the RJMCMC algorithm. Hence, the proposal distribution in Eq. (28) of merging MF l and m into MF n factors as

$$\begin{aligned} q(\text{merge}) &= q(\hat{\mathbf{c}}, \hat{\mathbf{z}}, \hat{w}_n, \hat{R}_n, \hat{\mathbf{\Sigma}}_{n,\{1\ldots6\}}|\mathbf{c}, w_{\{l,m\}}, R_{\{l,m\}}, \mathbf{\Sigma}_{\{l,m\}}, \mathbf{q}; \alpha, \gamma, \Delta, \nu) \\ &= q(\hat{\mathbf{c}}_{\{\mathbf{c}\in\{l,m\}\}}|\mathbf{c}) \prod_{i:\hat{c}_i=n} q(\hat{z}_i|w_l, R_l, \mathbf{\Sigma}_l, \mathbf{q}) q(\hat{R}_n|R_l, \hat{\mathbf{z}}, \hat{\mathbf{c}}, \mathbf{q}) q(\hat{\mathbf{\Sigma}}_{n,\{1\ldots6\}}|\hat{\mathbf{c}}, \hat{\mathbf{z}}, \mathbf{q}, \hat{R}_n; \Delta, \nu)\,. \end{aligned} \tag{29}$$

We restate the individual factors from Sec. 1.2 here for completeness:

$$q(\hat{\mathbf{c}}_{\{\mathbf{c}\in\{l,m\}\}}|\mathbf{c}) = \delta(\hat{\mathbf{c}}_{\{\mathbf{c}\in\{l,m\}\}} = n) \tag{30}$$

$$q(\hat{z}_i = j|w_l, R_l, \mathbf{\Sigma}_l, \mathbf{q}) \propto w_{lj}\, p(q_i; [M_l]_j, \Sigma_{lj}) \tag{31}$$

$$q(\hat{R}_n|R_l, \hat{\mathbf{z}}, \hat{\mathbf{c}}, \mathbf{q}) = \mathcal{N}(\hat{R}_n; \hat{R}_n^\star(R_l, \hat{\mathbf{z}}, \hat{\mathbf{c}}, \mathbf{q}), \Sigma_{\mathrm{so}(3)}) \tag{32}$$

$$q(\hat{\mathbf{\Sigma}}_{n,\{1\ldots6\}}|\hat{\mathbf{c}}, \hat{\mathbf{z}}, \mathbf{q}, \hat{R}_n; \Delta, \nu) = \prod_{j=1}^{6} p(\hat{\Sigma}_{nj}|\hat{\mathbf{z}}, \hat{\mathbf{c}}, \mathbf{q}, \hat{R}_n; \Delta, \nu)\,. \tag{33}$$

The split proposal distribution in Eq. (27) factors into

$$\begin{aligned} q(\text{split}) &= q(\mathbf{c}, \mathbf{z}, w_{\{l,m\}}, R_{\{l,m\}}, \mathbf{\Sigma}_{\{l,m\},\{1\ldots6\}}|\hat{\mathbf{c}}, \hat{w}_n, \hat{R}_n, \hat{\mathbf{\Sigma}}_n, \mathbf{q}; \alpha, \gamma, \Delta, \nu) = \\ &= q(\mathbf{c}_{\{\hat{\mathbf{c}}=n\}}|\hat{\mathbf{c}}; \alpha) \textstyle\prod_{i:\hat{c}_i=n} q(z_i|\hat{w}_n, \hat{R}_n, \hat{\mathbf{\Sigma}}_n, \mathbf{q}) q(R_l, R_m|\mathbf{z}, \mathbf{c}, \mathbf{q}, \hat{R}_n) q(\mathbf{\Sigma}_{\{l,m\},\{1\ldots6\}}|\mathbf{c}, \mathbf{z}, \mathbf{q}, R_{\{l,m\}}; \Delta, \nu)\,. \end{aligned} \tag{34}$$

The individual factors were introduced in Seq. 1.3 and are restated here for completeness:

$$q(\mathbf{c}_{\{\hat{\mathbf{c}}=n\}}|\hat{\mathbf{c}}; \alpha) = \mathrm{DirMult}\left(\mathbf{c}_{\{\hat{\mathbf{c}}=n\}}; \alpha_l, \alpha_m\right) \tag{35}$$

$$q(z_i = j|\hat{w}_n, \hat{R}_n, \hat{\mathbf{\Sigma}}_n, \mathbf{q}) \propto \hat{w}_{nj}\, p(q_i; [\hat{M}_n]_j, \hat{\Sigma}_{nj}) \tag{36}$$

$$q(R_l, R_m|\mathbf{z}, \mathbf{c}, \mathbf{q}, \hat{R}_n) = \mathcal{N}(R_l; R_l^\star(\hat{R}_n, \mathbf{z}, \mathbf{c}, \mathbf{q}), \Sigma_{\mathrm{so}(3)})\, \mathcal{N}(R_m; R_m^\star(\hat{R}_n, \mathbf{z}, \mathbf{c}, \mathbf{q}), \Sigma_{\mathrm{so}(3)}) \tag{37}$$

$$q(\mathbf{\Sigma}_{\{l,m\},\{1\ldots6\}}|\mathbf{c}, \mathbf{z}, \mathbf{q}, R_{\{l,m\}}; \Delta, \nu) = \prod_{j=1}^{6} p(\Sigma_{lj}|\mathbf{z}, \mathbf{c}, \mathbf{q}, R_l; \Delta, \nu) p(\Sigma_{mj}|\mathbf{z}, \mathbf{c}, \mathbf{q}, R_m; \Delta, \nu)\,. \tag{38}$$

All in all, the ratio for the right side of the $\min(1, \cdot)$ in the acceptance probability for a merge move in Eq. (14) is:

$$\begin{aligned} &\frac{8\pi^2\left(\prod_{i=1}^{N} p(q_i|\hat{c}_i, \hat{z}_i, \hat{\mathbf{R}}, \hat{\mathbf{\Sigma}})\right) p(\hat{\mathbf{c}};\alpha) \prod_{k=1}^{\hat{K}} p(\hat{\mathbf{z}}_{\{\hat{\mathbf{c}}=k\}}|\hat{\mathbf{c}};\gamma) \prod_{j=1}^{6} p(\hat{\Sigma}_{kj};\Delta,\nu)}{\left(\prod_{i=1}^{N} p(q_i|c_i, z_i, \mathbf{R}, \mathbf{\Sigma})\right) p(\mathbf{c};\alpha) \prod_{k=1}^{K} p(\mathbf{z}_{\{\mathbf{c}=k\}}|\mathbf{c};\gamma) \prod_{j=1}^{6} p(\Sigma_{kj};\Delta,\nu)} \frac{q(\mathbf{c},\mathbf{z},w_{\{l,m\}},R_{\{l,m\}},\mathbf{\Sigma}_{\{l,m\},\{1\ldots6\}}|\hat{\mathbf{c}},\hat{w}_n,\hat{R}_n,\hat{\mathbf{\Sigma}}_n,\mathbf{q};\alpha,\gamma,\Delta,\nu)}{q(\hat{\mathbf{c}},\hat{\mathbf{z}},\hat{w}_n,\hat{R}_n,\hat{\mathbf{\Sigma}}_{n,\{1\ldots6\}}|\mathbf{c},w_{\{l,m\}},R_{\{l,m\}},\mathbf{\Sigma}_{\{l,m\}},\mathbf{q};\alpha,\gamma,\Delta,\nu)} = \\ = \; & 8\pi^2 \prod_{i=1}^{N} \frac{p(q_i|\hat{c}_i, \hat{z}_i, \hat{\mathbf{R}}, \hat{\mathbf{\Sigma}})}{p(q_i|c_i, z_i, \mathbf{R}, \mathbf{\Sigma})} \prod_{i:\hat{c}_i=n} \frac{q(z_i|\hat{w}_n, \hat{R}_n, \hat{\mathbf{\Sigma}}_n, \mathbf{q})}{q(\hat{z}_i|w_l, R_l, \mathbf{\Sigma}_l, \mathbf{q})} \prod_{j=1}^{6} \frac{p(\Sigma_{lj}|\mathbf{z},\mathbf{c},\mathbf{q},R_l;\Delta,\nu) p(\Sigma_{mj}|\mathbf{z},\mathbf{c},\mathbf{q},R_m;\Delta,\nu) p(\hat{\Sigma}_{nj};\Delta,\nu)}{p(\Sigma_{lj};\Delta,\nu) p(\Sigma_{mj};\Delta,\nu) p(\hat{\Sigma}_{nj}|\hat{\mathbf{z}},\hat{\mathbf{c}},\mathbf{q},\hat{R}_n;\Delta,\nu)} \\ & \frac{\mathrm{DirMult}(\hat{\mathbf{c}};\alpha)\, \mathrm{DirMult}(\hat{\mathbf{z}}_{\{\hat{\mathbf{c}}=n\}};\gamma)\, \mathrm{DirMult}\left(\mathbf{c}_{\{\hat{\mathbf{c}}=n\}};\alpha_l,\alpha_m\right)}{\mathrm{DirMult}(\mathbf{c};\alpha)\, \mathrm{DirMult}(\mathbf{z}_{\{\mathbf{c}=l\}};\gamma)\, \mathrm{DirMult}(\mathbf{z}_{\{\mathbf{c}=m\}};\gamma)\, \delta(\hat{\mathbf{c}}_{\{\mathbf{c}\in\{l,m\}\}} - n)} \frac{\mathcal{N}(R_l; R_l^\star(\hat{R}_n, \mathbf{z}, \mathbf{c}, \mathbf{q}), \Sigma_{\mathrm{so}(3)})\, \mathcal{N}(R_m; R_m^\star(\hat{R}_n, \mathbf{z}, \mathbf{c}, \mathbf{q}), \Sigma_{\mathrm{so}(3)})}{\mathcal{N}(\hat{R}_n; \hat{R}_n^\star(R_l, \hat{\mathbf{z}}, \hat{\mathbf{c}}, \mathbf{q}), \Sigma_{\mathrm{so}(3)})} \end{aligned} \tag{39}$$

Plugging Eq. (39) into Eq. (14), we can compute the acceptance probability for a merge of two MFs l and m into MF n. For a split of MF l into MFs m and n the acceptance probability is computed using the inverse ratio of Eq. (39).

2. Normal Distribution over Rotation Matrices

A matrix $R \in \mathbb{R}^{3\times3}$ is called a *rotation matrix* if it is an element of $\mathrm{SO}(3)$, the Special Orthogonal group; namely, $R^T R = \mathrm{I}$ and $\det(R) = 1$. Probability distributions over rotation matrices can be defined by exploiting the manifold structure of $\mathrm{SO}(3)$ (*e.g.*, [2, 6]). In particular, one well-known way to construct an analog of a Gaussian distribution over this nonlinear space utilizes the linearity of the tangent spaces.

Let $\mathrm{Log}_{R_\mu}(R) : \mathrm{SO}(3) \to T_{R_\mu}\mathrm{SO}(3)$ denote the Riemannian logarithm map of R into the tangent space $T_{R_\mu}\mathrm{SO}(3)$ around R_μ:

$$\theta = \arccos\left(\frac{\mathrm{trace}(R_\mu^T R) - 1}{2}\right)\,; \tag{40}$$

$$\mathrm{Log}_{R_\mu}(R) = R_\mu \left(\frac{\theta}{2\sin(\theta)}\left(R_\mu^T R - R^T R_\mu\right)\right)\,. \tag{41}$$

While $\mathrm{Log}_{R_\mu}(R)$ is in $T_{R_\mu}\mathrm{SO}(3)$, the matrix $W = R_\mu^T \mathrm{Log}_{R_\mu}(R)$ is an element of the Lie Algebra $\mathrm{so}(3)$; *i.e.*, W is skew-symmetric. The *vee operator* $^\vee$ [2] extracts the unique elements of W into a vector w: $W^\vee = w = [-W_{23}; W_{13}; -W_{12}] \in \mathbb{R}^3$. The Riemannian logarithm map in conjunction with the vee operator allows us to define a normal distribution with mean rotation R_μ and covariance $\Sigma_{\mathrm{so}(3)}$ in the tangent space $T_{R_\mu}\mathrm{SO}(3)$:

$$p(R; R_\mu, \Sigma_{\mathrm{so}(3)}) = \mathcal{N}((R_\mu^T \mathrm{Log}_{R_\mu}(R))^\vee; 0, \Sigma_{\mathrm{so}(3)})\,, \tag{42}$$

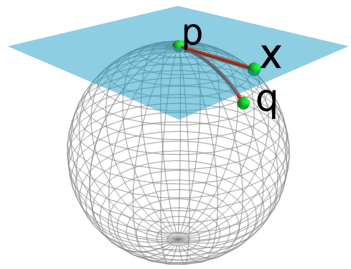

Figure 1: The unit sphere S^2 in 3 dimensions. The blue plane on the sphere illustrates T_pS^2, the tangent space to S^2 at point p. A tangent vector $x \in T_pS^2$ is mapped to $q \in S^2$ via Exp_p, the Riemannian exponential map with respect to p.

where $\Sigma_{\mathrm{so}(3)} \in \mathbb{R}^{3\times 3}$.

In order to sample from the distribution in Eq. (42), we sample $w = W^\vee \sim \mathcal{N}(0, \Sigma_{\mathrm{so}(3)})$ and map it back from the tangent space $T_{R_\mu}\mathrm{SO}(3)$ to $\mathrm{SO}(3)$ using the Riemannian exponential map $\mathrm{Exp}_{R_\mu} : T_{R_\mu}\mathrm{SO}(3) \to \mathrm{SO}(3)$:

$$\theta = ||w||_2 \tag{43}$$

$$\mathrm{Exp}_{R_\mu}(W) = R_\mu \left(\mathrm{I} + \frac{\sin(\theta)}{\theta} W + \frac{1-\cos(\theta)}{\theta^2} W^2 \right) \tag{44}$$

For further details on the manifold of rotations $\mathrm{SO}(3)$, the log and exp maps we have briefly reviewed here, as well as the relation to the Lie Algebra $\mathrm{so}(3)$, refer to [2, 3].

3. Geodesic Distances and the Riemannian Exponential/Log map for the Unit Sphere S^2

Let p and q be two points on the unit sphere in 3D S^2, and let T_pS^2 denote the tangent space to S^2 at point p; namely,

$$p^T p = q^T q = 1 \tag{45}$$

$$T_pS^2 = \{x : x \in \mathbb{R}^3 \; ; \; x^T p = 0\} \, . \tag{46}$$

Note that while S^2 is a nonlinear manifold, T_pS^2 is a 2-dimensional linear space as depicted in Fig. 1. It can be shown [1, 3] that the geodesic distance between p and q is given by the angle between p and q

$$d_G(p, q) = \arccos(p^T q) \, . \tag{47}$$

Furthermore, the Riemannian exponential map $\mathrm{Exp}_p : T_pS^2 \to S^2$ maps a point x in the tangent space T_pS^2 around p onto the sphere S^2:

$$x \mapsto p\cos(||x||) + \tfrac{x}{||x||}\sin(||x||) \, . \tag{48}$$

The inverse of Exp_p, the Riemannian logarithm map $\mathrm{Log}_p : S^2/\{-p\} \to T_pS^2$ can be computed as:

$$q \mapsto (q - p\cos\theta)\tfrac{\theta}{\sin\theta} \, , \tag{49}$$

where $\theta = d_G(p, q)$.

In other words, the geodesic distance between two unit normals is the angle between them, Exp_p maps T_pS^2 onto S^2, and Log_p, whose action can be thought of as a linearization of S^2, is defined over the entire sphere *except* the antipodal point $-p$. Note that Exp_p and Log_p depend on p. For further details and an introduction to Riemannian geometry see [3].

References

[1] P. Absil, R. Mahony, and R. Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009. 6

[2] G. S. Chirikjian. *Stochastic Models, Information Theory, and Lie Groups, Volume 2: Analytic Methods and Modern Applications*, volume 2. Springer, 2011. 2, 4, 5, 6

[3] M. P. do Carmo. *Riemannian Geometry*. Birkhäuser Verlag, Boston, MA, 1992. 6

[4] P. J. Green. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82(4):711–732, 1995. 1

[5] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970. 1

[6] X. Pennec. Probabilities and statistics on Riemannian manifolds: Basic tools for geometric measurements. In *NSIP*, pages 194–198, 1999. 5

[7] S. Richardson and P. J. Green. On bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society: series B (statistical methodology)*, 59(4):731–792, 1997. 1

[8] J. Straub, G. Rosman, O. Freifeld, J. J. Leonard, and J. W. Fisher III. A mixture of manhattan frames: Beyond the manhattan world. In *CVPR*, 2014. 1