SBM for point set registration

Giulia Marchello[1], Marco Corneli[2,3], Davide Adamo[2,3] & Charles Bouveyron[2]

[1] *Inria PreMeDICaL team, Idesp, Université de Montpellier, France.*

[2] *Université Côte d'Azur, Inria, CNRS, Laboratoire J.A.Dieudonné, Maasai team, Nice, France.*

[3] *Université Côte d'Azur, Laboratoire CEPAM, Nice, France.*

Résumé. L'enregistrement de nuages de points est une tâche fondamentale en vision avec des applications dans la recherche d'images 3D, la segmentation et la reconnaissance de formes. Cet article aborde les défis de l'enregistrement d'ensembles de points, en tenant compte de facteurs tels que les transformations spatiales non rigides, la grande dimensionalité, le bruit et les valeurs aberrantes. L'accent est mis sur les méthodes basées sur des probabilitées, en exploitant le modèle à blocs stochastiques (SBM). L'approche proposée consiste à représenter les nuages de points sous forme de graphes et à introduire une variable latente pour le regroupement. Pour améliorer la parcimonie et l'efficacité computationnelle, le modèle intègre une distribution normale vec excès de zéros, se concentrant uniquement sur les entrées non nulles en dessous d'un seuil spécifié. L'article décrit la distribution conjointe et présente un algorithme d'inférence variationnelle pour l'estimation des paramètres. La méthodologie offre un cadre probabiliste pour un enregistrement robuste d'ensembles de points, démontrant son potentiel dans des scénarios complexes avec des données de grande dimensionalité. En tant que document de travail, les étapes suivantes concerneront des exemples numériques sur des ensembles de données simulés et réels.

Mots-clés. Enregistrement de Nuages de Points, Modèle à Blocs Stochastiques, Distribution Normale avec excès de zéros, Inférence Variationnelle, Transformations Spatiales.

Abstract. The registration of point clouds is a fundamental task in computer vision with applications in 3D image retrieval, segmentation, and shape recognition. This paper addresses the challenges of point set registration, considering factors such as nonrigid spatial transformations, high dimensionality, noise, and outliers. The focus is on probability-based methods, specifically leveraging the Stochastic Block Model (SBM). The proposed approach involves representing point clouds as graphs and introducing a latent variable for clustering. The proposed approach involves representing point clouds through graphs, introducing a latent variable for clustering. To enhance sparsity and computational efficiency, the model incorporates a Zero-Inflated Normal distribution, focusing solely on non-zero entries below a specified threshold. The paper outlines the joint distribution and presents a variational inference algorithm for parameter estimation. The methodology provides a probabilistic framework for robust point set registration, demonstrating its potential in complex scenarios with high-dimensional data. Being a working paper, the following steps will concern numerical examples on simulated and real dataset.

Keywords. Point Cloud Registration, Stochastic Block Model, Zero-Inflated Normal Distribution, Variational Inference, Spatial Transformations.

1 Introduction

Registration of point clouds is crucial in various computer vision applications, playing a pivotal role in tasks such as 3D image retrieval, segmentation, and shape recognition. The main goal of point set registration consists in establishing correspondences between two sets of points and determining transformations that maps one point set onto another. Point set registration is a challenging task due to several factors, such as the presence of an unknown nonrigid spatial transformation, the high dimensionality of point sets, potential noise, and the existence of outliers. The transformation considered in point set registration typically falls in two categories: rigid or nonrigid. A rigid transformation is constrained to translation, rotation, and scaling. On the other hand, nonrigid transformations encompass a broader range of alterations, such as stretching and skewing. Figure 1 shows an example of point set registration with a 2D toy example. In this picture, the second image is visibly rotated by 90 degrees and it and contains an additional point denoted by a question mark compared to the first image. This discrepancy adds complexity to the matching task. Extending this concept to more intricate scenarios involving high-dimensional data, noise, outliers, and 3D images, it becomes evident that point sets registration can quickly become a challenging task.

Regarding the pairwise point set registration, depending on the modeling assumptions, several approaches have been proposed in the literature. They can be classified into distance-based methods (Zhang, 1994; Zhou and De la Torre, 2015), filter-based methods (Zhu et al., 2018; Li et al., 2016) and probability-based methods (Myronenko and Song, 2010; Zhou et al., 2014). It has been observed that probability-based methods tend to outperform other approaches. However, it is worth noting that the probability-based methods come at a higher computational cost in contrast to distance-based and filter-based methods (Zhu et al., 2019).

This section aims to formulate a probability-based point set registration method, designed to effectively address rigid transformations in the context of 3D point clouds.

1.1 Our contribution

Our approach would be to use a probabilistic method as well. Specifically, we plan on making use of the Stochastic Block Model instead of Gaussian Mixture Models (GMMs) in order to model the points of a source cloud with those of a target cloud. In more detail, we represent a point cloud with N points via a graph with N nodes, thus leading to a (weighted) adjacency $N \times N$ matrix, denoted by X. The way the nodes are connected to each other and hence the nature of the graph is of course crucial and several approaches are possible. The idea is indeed to use a an ϵ-neighborhood graph where Θ_{ij} being either the Euclidean distance between to neighbor nodes or zero, with: $i, j = 1, \ldots, N$:

$$\Theta_{ij} = \begin{cases} d(i,j) & \text{if } d(i,j) < \epsilon, \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Also, one might contemplate using a fully connected graph, where X_{ij} denotes the geodesic distance between points i and j. However, it is imperative to acknowledge that the compu-

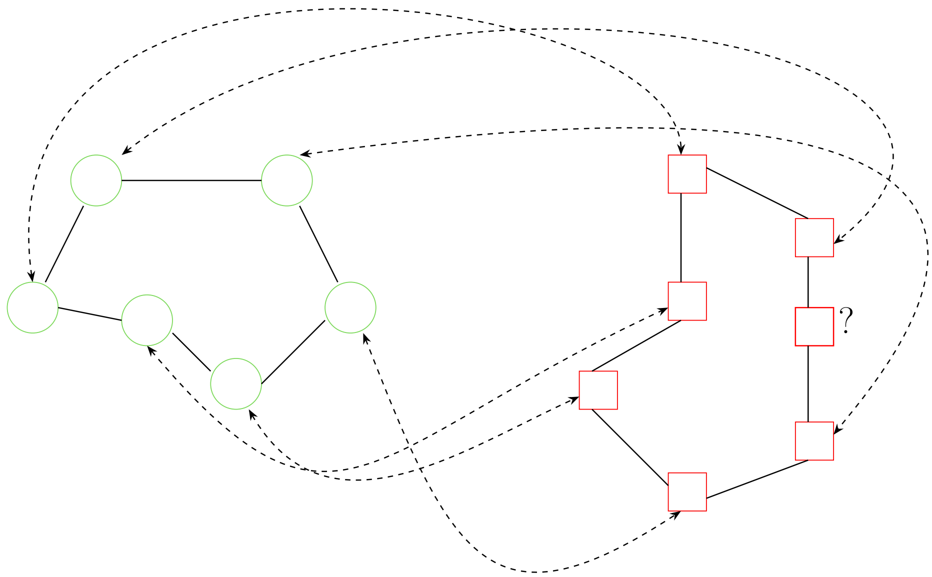

Figure 1: A toy example of the point set registration problem.

tational cost associated with this approach tends to escalate proportionally with the growing number of points.

For simplicity, we focus on two point clouds $\mathcal{X}$ and $\mathcal{Y}$ represented by two matrices of dimension $N \times 3$ and $Q \times 3$, respectively, with $N > Q$. We compute their distance adjacency matrices following Eq.(1), obtaining $\Theta^{\mathcal{X}}$ and $\Theta^{\mathcal{Y}}$, respectively.

Now, let Z be a latent matrix such that: $Z := \{z_{iq}\}_{i \in 1,\dots,N, q \in 1,\dots,Q}$. This matrix represents the clustering of point of $\mathcal{X}$ into Q groups, where point i belongs to cluster q if $z_{iq} = 1$, 0 otherwise. Moreover, the rows of Z are assumed to be independently distributed according to a multinomial distributions:

$$p(Z|\alpha) = \prod_{i=1}^{N} \prod_{q=1}^{Q} \alpha_q^{z_{iq}},$$

where $\alpha_q = \mathbb{P}\{z_{iq} = 1\}$ and $\sum_{q=1}^{Q} \alpha_q = 1$.

We further assume that, conditionally to Z, $\Theta^{\mathcal{X}}$ follows a Zero-Inflated Normal distribution, such that:

$$\Theta_{ij}^{\mathcal{X}} | Z_i Z_j \sim \mathcal{ZI}N(\Theta_{Z_i Z_j}^{\mathcal{Y}}; \sigma_{Z_i Z_j}^2). \quad (2)$$

Being a mixture between a chosen distribution and a Dirac mass at zero, the Zero-Inflated distribution is used to account for a high sparsity in the data and can be formally written as:

$$\begin{cases} \Theta^{\mathcal{X}}_{ij}|Z_i, Z_j \sim \delta_0(\Theta^{\mathcal{X}}_{ij}) & \text{with probability } \pi_{Z_iZ_j} \\ \Theta^{\mathcal{X}}_{ij}|Z_i, Z_j \sim N(\Theta^{\mathcal{X}}_{ij}; \Theta^{\mathcal{Y}}_{Z_iZ_j}; \sigma^2_{Z_iZ_j}) & \text{with probability } 1-\pi_{Z_iZ_j}. \end{cases} \tag{3}$$

where $\delta_0(\cdot)$ is the Dirac function in 0 and $\Theta^{\mathcal{Y}}_{Z_iZ_j}, \sigma^2_{Z_iZ_j}$ are the block-dependent parameters of the Normal distribution.

Then, to model the sparsity, we rewrite Eq. (3) by introducing a hidden random matrix, $A \in \{0,1\}^{N\times N}$, where:

$$A_{ij}|Z_iZ_j \sim \mathcal{B}(\pi_{Z_iZ_j}),$$

with $\mathcal{B}(\pi)$ denoting the Bernoulli probability mass function of parameter π and such that

$$\begin{aligned} &A_{ij} = 1 \Rightarrow \Theta^{\mathcal{X}}_{ij}|Z_i, Z_j \sim \delta_0(\Theta^{\mathcal{X}}_{ij}) \\ &A_{ij} = 0 \Rightarrow \Theta^{\mathcal{X}}_{ij}|Z_i, Z_j \sim N(\Theta^{\mathcal{X}}_{ij}; \Theta^{\mathcal{Y}}_{Z_iZ_j}; \sigma^2_{Z_iZ_j}). \end{aligned} \tag{4}$$

2 The joint distribution

The model described so far has a set of parameters denoted by $\theta := (\Theta^{\mathcal{Y}}, \Sigma, \alpha, \pi)$, with $\Theta^{\mathcal{Y}} =: \{\Theta^{\mathcal{Y}}_{ql}\}_{q,l\leq Q}$ and $\Sigma := \{\sigma^2_{ql}\}_{q,l}$ and a set of latent variables: $\{Z, A\}$.
As a first move, we can compute the likelihood of the complete data:

$$p(\Theta^{\mathcal{X}}, A, Z|\theta) = p(\Theta^{\mathcal{X}}|A, Z, \Theta^{\mathcal{Y}}, \Sigma)p(A|Z, \pi)p(Z|\alpha), \tag{5}$$

where:

$$p(\Theta^{\mathcal{X}}|A, Z, \Theta^{\mathcal{Y}}, \Sigma) = \prod_{j>i}^{N} \mathbf{1}^{A_{ij}}_{\{\Theta^{\mathcal{X}}_{ij}=0\}} \left\{ \left(\mathcal{N}(\Theta^{\mathcal{X}}_{ij}; \Theta^{\mathcal{Y}}_{Z_iZ_j}; \sigma^2_{Z_iZ_j}) \right)^{(1-A_{ij})} \right\},$$

$$p(A|Z, \pi) = \prod_{j>i}^{N} \pi^{A_{ij}}_{Z_iZ_j} \left(1 - \pi_{Z_iZ_j}\right)^{(1-A_{ij})},$$

$$p(Z|\alpha) = \prod_{i=1}^{N} \prod_{q=1}^{Q} \alpha_q^{Z_{iq}},$$

where the shorthand notation $\sum_{j>i}^{N} := \sum_{i=1}^{N} \sum_{j>i}^{N}$ was adopted. By looking at the first of the above equations, it is immediately clear that if $A_{ij} = 1$ and $\Theta^{\mathcal{X}}_{ij} \neq 0$ the whole likelihood goes to zero. On the contrary, we assume that $\mathbf{1}^{A_{ij}}_{\{\Theta^{\mathcal{X}}_{ij}=0\}}$ takes value one when both A_{ij} and the indicator function are zero. So, in order to have a positive likelihood, we must assume that we never observe $A_{ij} = 1$ and $\Theta^{\mathcal{X}}_{ij} \neq 0$ and, as a consequence, we could replace the indicator function with 1. However we let it in place since it will provide us with an interesting intuition, later. Hence, we can write the log-likelihood of the complete data as follows:

$$\log p(\Theta^{\mathcal{X}}, Z, A|\theta) = \frac{1}{2}\sum_{j\neq i}^{N}\sum_{q,\ell}^{Q}\Big[A_{ij}Z_{iq}Z_{j\ell}\log\big(\pi_{ql}\mathbf{1}_{\{\Theta^{\mathcal{X}}_{ij}=0\}}\big)+ \\ + (1-A_{ij})Z_{iq}Z_{j\ell}\log\big((1-\pi_{ql})\mathcal{N}(\Theta^{\mathcal{X}}_{ij};\Theta^{\mathcal{Y}}_{ql},\sigma^2_{ql})\big)\Big]+ \\ + \sum_{i=1}^{N}\sum_{q=1}^{Q} Z_{iq}\log\alpha_q. \tag{6}$$

3 The inference

3.1 Variational assumptions

Since we cannot compute the posterior distribution, $p(A, Z|\Theta^{\mathcal{X}}, \theta)$, we rely on a variational procedure which optimizes a lower bound of the likelihood. Let us thus introduce a variational distribution $q(.)$ in order to decompose the likelihood as follows:

$$\log p(\Theta^{\mathcal{X}}|\theta) = \mathcal{L}(q,\theta) + KL(q \parallel p_{A,Z}),$$

where $\mathcal{L}$ denotes a lower bound defined as

$$\mathcal{L}(q,\theta) = \sum_{A}\sum_{Z} q(A,Z)\log\frac{p(\Theta^{\mathcal{X}}, A, Z|\theta)}{q(A,Z)} \\ = E_{q(A,Z)}\Big[\log\frac{p(\Theta^{\mathcal{X}}, A, Z|\theta)}{q(A,Z)}\Big] \\ = E_{q(A,Z)}[\log(p(\Theta^{\mathcal{X}}, A, Z|\theta)] - E_{q(A,Z)}[\log(q(A,Z))], \tag{7}$$

and KL indicates the Kullaback-Liebler divergence between the true and the approximate posterior $p_{A,Z} := p(A, Z|\Theta^{\mathcal{X}}, \theta)$:

$$KL(q \parallel p_{A,Z}) = -\sum_{A}\sum_{Z} q(A,Z)\log\frac{p(A,Z|\Theta^{\mathcal{X}},\theta)}{q(A,Z)}.$$

Now, the objective is to find a distribution $q(\cdot)$ that maximizes the lower bound $\mathcal{L}(q,\theta)$. In order to allow the optimization of $\mathcal{L}(q,\theta)$, we further assume that $q(A,Z)$ factorizes as follows:

$$q(A,Z) = q(A)q(Z) = \prod_{j\neq i}^{N} q(A_{ij})\prod_{i=1}^{N} q(Z_i) \\ = \prod_{j\neq i}^{N}\delta_{ij}^{A_{ij}}(1-\delta_{ij})^{(1-A_{ij})}\prod_{i=1}^{N}\prod_{q}^{Q}\tau_{iq}^{Z_{iq}},$$

Where δ_{ij} and τ_{iq} indicate the variational parameters of A_{ij} and τ_{iq}, respectively (see Section 3.2 for details).

3.2 VE-Step

The VE-step of the VEM algorithm aims at maximizing the lower bound in Eq. (7) with respect to the variational distribution $q(\cdot)$ while keeping θ fixed. Following **?**, we derive the update equations for the factors $q(A)$ and $q(Z)$, such that the log of the optimized factors are given by:

$$\log q^*(A) = E_{q(Z)}[\log p(\Theta^{\mathcal{X}}, A, Z|\theta)], \tag{8}$$

$$\log q^*(Z) = E_{q(A)}[\log p(\Theta^{\mathcal{X}}, A, Z|\theta)], \tag{9}$$

3.2.1 Optimization of the factor q(A)

Let us consider the derivation of the update equation for the factor $q(A)$. The sequential update for the factor $q(A)$ can be computed through the logarithm of the optimized factor, where all the terms that do not depend on A are absorbed in a constant.

Proposition 1. *Denoting by δ_{ij} the variational success probability for A_{ij}, the optimal update of $q(A)$ is:*

$$\delta_{ij} = \frac{\exp(R_{ij})}{1+\exp(R_{ij})}, \tag{10}$$

with:

$$R_{ij} := \log \mathbf{1}_{\{\Theta^{\mathcal{X}}_{ij}=0\}} + \frac{1}{2}\sum_{q,\ell}^{Q} \tau_{iq}\tau_{j\ell}\left(\log\left(\frac{\pi_{q\ell}}{1-\pi_{q\ell}}\right) - \log \mathcal{N}(\Theta^{\mathcal{X}}_{ij}; \Theta^{\mathcal{Y}}_{ql}; \sigma^2_{ql})\right). \tag{11}$$

Proof.

$$\begin{aligned}
\log q^*(A) &= E_{q(Z)}[\log p(\Theta^{\mathcal{X}}, A, Z|\theta)] \\
&= \frac{1}{2}\sum_{j\neq i}^{N}\sum_{q,\ell}^{Q}\Big[A_{ij}\tau_{iq}\tau_{j\ell}\log\left(\pi_{q\ell}\mathbf{1}_{\{\Theta^{\mathcal{X}}_{ij}=0\}}\right)+ \\
&+ (1-A_{ij})\tau_{iq}\tau_{j\ell}\log\left((1-\pi_{q\ell})\mathcal{N}(\Theta^{\mathcal{X}}_{ij}; \Theta^{\mathcal{Y}}_{q\ell}; \sigma^2_{q\ell})\right)\Big] + C \\
&= \frac{1}{2}\sum_{j\neq i}^{N}\sum_{q,\ell}^{Q}\Big[A_{ij}\tau_{iq}\tau_{j\ell}\log\left(\pi_{q\ell}\mathbf{1}_{\{\Theta^{\mathcal{X}}_{ij}=0\}}\right)+ \\
&- A_{ij}\tau_{iq}\tau_{j\ell}\log\left((1-\pi_{q\ell})\mathcal{N}(\Theta^{\mathcal{X}}_{ij}; \Theta^{\mathcal{Y}}_{q\ell}; \sigma^2_{q\ell})\right)\Big] + C \\
&= \frac{1}{2}\sum_{j\neq i}^{N} A_{ij}\left[\sum_{q,\ell}^{Q}\tau_{iq}\tau_{j\ell}\Big(\log\left(\frac{\pi_{q\ell}}{1-\pi_{q\ell}}\right) + \log(\mathbf{1}_{\{\Theta^{\mathcal{X}}_{ij}=0\}})+\right. \\
&\left. - \log(\mathcal{N}(\Theta^{\mathcal{X}}_{ij}; \Theta^{\mathcal{Y}}_{q\ell}; \sigma^2_{q\ell}))\Big)\right] + C.
\end{aligned} \tag{12}$$

□

We can then recognize the functional form of a Bernoulli distribution:

$$\begin{aligned}\log q^*(A) &= \sum_{j\neq i}^{N} A_{ij}\log\delta_{ij} + (1-A_{ij})\log(1-\delta_{ij}), \\ &= \sum_{j\neq i}^{N} A_{ij}\log\left(\frac{\delta_{ij}}{1-\delta_{ij}}\right) + C,\end{aligned} \tag{13}$$

where δ_{ij} is

$$\delta_{ij} = \frac{\exp(R_{ij})}{1+\exp(R_{ij})}, \tag{14}$$

with R_{ij} defined in Eq. (11).

Note that, although everything in the above equations is well defined in force of the assumptions we made (in particular one never has $\log(0)$ unless $A_{ij} = 0$) formally, when $\Theta^{\mathcal{X}}_{ij} \neq 0$, $R_{ij} = -\infty$ and $\delta_{ij} = 0$, which makes sense: non-null distances in $\Theta^{\mathcal{X}}_{ij}$ come from a Normal distribution distribution with probability one (see Eq. 4).

3.2.2 Optimization of the factor q(Z)

Let us consider the derivation of the update equation for the factor $q(Z)$. The sequential update for the factor $q(Z)$ can be computed through the log of the optimized factor.

Proposition 2. *Denoting by τ_{iq} the variational success probability for Z_{iq} the optimal update of $q(Z)$ is:*

$$\tau_{iq} = \frac{r_{iq}}{\sum_{\ell=1}^{Q} r_{i\ell}}, \tag{15}$$

with:

$$r_{iq} \propto \alpha_q \exp\left(\sum_{\substack{j=1\\ j\neq i}}^{N}\sum_{\ell=1}^{Q}\tau_{j\ell}\left[\delta_{ij}\log\left(\pi_{q\ell}\mathbf{1}_{\{\Theta^{\mathcal{X}}_{ij}=0\}}\right) + (1-\delta_{ij})\log\left((1-\pi_{q\ell})\mathcal{N}(\Theta^{\mathcal{X}}_{ij};\Theta^{\mathcal{Y}}_{q\ell},\sigma^2_{q\ell})\right)\right]\right) \tag{16}$$

Proof. First, we know that

$$\log q^*(Z_i) = E_{A,Z_{/i}\sim q}[\log p(\Theta^{\mathcal{X}}, A, Z|\theta))]$$

where the subscript $A, Z_{/i} \sim q$ means that the expectation is taken with respect to all the latent variables following the variational distribution *except* Z_i. Thence we can isolate the terms in $\log p(\Theta^{\mathcal{X}}, A, Z|\theta))$ depending on Z_i and take the expectation

$$\begin{aligned}\log q^*(Z_i) = &\sum_{j\neq i}^{N}\sum_{q,\ell}^{Q} Z_{iq}\left(\delta_{ij}\tau_{j\ell}\log\left(\pi_{q\ell}\mathbf{1}_{\{\Theta^{\mathcal{X}}_{ij}=0\}}\right) + (1-\delta_{ij})\tau_{j\ell}\log\left((1-\pi_{q\ell})\mathcal{N}(\Theta^{\mathcal{X}}_{ij};\Theta^{\mathcal{Y}}_{q\ell},\sigma^2_{q\ell})\right)\right) \\ &+ \sum_{q=1}^{Q} Z_{iq}\log\alpha_q + C\end{aligned}$$

where C regroups all the terms not depending on Z_i and we stress that here $\sum_{j\neq i}^N$ is $\sum_{\substack{j=1,\\ j\neq i}}^N$, with i being fixed. The above equation factorizes as follows

$$\log q^*(Z_i) = \sum_{q=1}^{Q} Z_{iq} \log r_{iq} + C \tag{17}$$

where

$$\log r_{iq} := \sum_{j\neq i}^{N} \sum_{\ell=1}^{Q} \tau_{j\ell} \left(\delta_{ij} \log\left(\pi_{q\ell} \mathbf{1}_{\{\Theta_{ij}^{\mathcal{X}}=0\}}\right) + (1-\delta_{ij}) \log\left((1-\pi_{q\ell}) \mathcal{N}(\Theta_{ij}^{\mathcal{X}}; \Theta_{q\ell}^{\mathcal{Y}}, \sigma_{q\ell}^2)\right)\right) + \log \alpha_q.$$

We recognize the log-likelihood of a multinomial distribution in Eq. (17) and taking the exponential on both sides of the above equation the proposition is proven. □

3.3 M-Step

In order to obtain the updating of the parameter set θ, the objective of the M-Step is the maximization of the lower bound $L(q;\theta)$ with respect to α and π. It is worth noticing that the parameter $\Theta_{q\ell}^{\mathcal{Y}}$ is observed (see Section 1.1) and that at this stage we fix the variance of the Normal distribution, Σ.
The final expression of the variational lower bound can be obtained by developing Eq. (7) as follows:

$$\begin{aligned}
\mathcal{L}(q,\theta) &= E_{q(A,Z)}\Big[\log p(X,A,Z|\theta)\Big] - E_{q(A,Z)}\Big[\log q(A,Z)\Big] \\
&= \frac{1}{2} \sum_{j\neq i}^{N} \sum_{q,\ell}^{Q} \left[\delta_{ij}\tau_{iq}\tau_{j\ell} \log(\pi_{q\ell} \mathbf{1}_{\{\Theta_{ij}^{\mathcal{X}}=0\}}) + (1-\delta_{ij})\tau_{iq}\tau_{j\ell} \log(1-\pi_{q\ell}) \mathcal{N}(\Theta_{ij}^{\mathcal{X}}; \Theta_{q\ell}^{\mathcal{Y}}, \sigma_{q\ell}^2) \right] + \\
&+ \sum_{i\neq 1}^{N} \sum_{q=1}^{Q} \tau_{iq} \log \alpha_q - \sum_{i\neq 1}^{N} \sum_{q=1}^{Q} \tau_{iq} \log \tau_{iq} - \sum_{j\neq i}^{N} \Big[\delta_{ij} \log(\delta_{ij}) + (1-\delta_{ij}) \log(1-\delta_{ij})\Big]
\end{aligned} \tag{18}$$

3.3.1 Update of π:

Here our goal is to derive the update of the block-dependent sparsity parameter, $\pi_{q\ell}$. The variational distribution $q(A,Z)$ is kept fixed, while the lower bound in Eq.(18) is maximized with respect to $\pi_{q\ell}$, to obtain its update,$\hat{\pi}_{q\ell}$.

Proposition 3. *The updating formula of π is obtained by maximizing $L(q;\theta)$ with respect to the parameter. Following algebraic manipulations, it can be expressed as:*

$$\hat{\pi}_{q\ell} := \frac{\sum_{j\neq i}^{N} \delta_{ij}\tau_{iq}\tau_{j\ell}}{\sum_{j\neq i}^{N} \left[\tau_{iq}\tau_{j\ell} \mathcal{N}(\Theta_{ij}^{\mathcal{X}}; \Theta_{q\ell}^{\mathcal{Y}}, \sigma_{q\ell}^2) + \delta_{ij}\tau_{iq}\tau_{j\ell} - \delta_{ij}\tau_{iq}\tau_{j\ell} \mathcal{N}(\Theta_{ij}^{\mathcal{X}}; \Theta_{q\ell}^{\mathcal{Y}}, \sigma_{q\ell}^2) \right]} \tag{19}$$

Proof.

$$
\begin{aligned}
\frac{\partial \mathcal{L}(q,\theta)}{\partial \pi_{q\ell}} &= \frac{1}{2}\sum_{j\neq i}^{N}\left[\frac{\delta_{ij}\tau_{iq}\tau_{j\ell}}{\pi_{q\ell}} - \frac{(1-\delta_{ij})\tau_{iq}\tau_{j\ell}\mathcal{N}(\Theta^{\mathcal{X}}_{ij};\Theta^{\mathcal{Y}}_{q\ell},\sigma^2_{q\ell})}{1-\pi_{q\ell}}\right] = 0 \\
&= \sum_{j\neq i}^{N}\Bigg[\delta_{ij}\tau_{iq}\tau_{j\ell} - \delta_{ij}\tau_{iq}\tau_{j\ell}\pi_{q\ell} - \pi_{q\ell}\tau_{iq}\tau_{j\ell}\mathcal{N}(\Theta^{\mathcal{X}}_{ij};\Theta^{\mathcal{Y}}_{q\ell},\sigma^2_{q\ell}) + \\
&+ \delta_{ij}\tau_{iq}\tau_{j\ell}\pi_{q\ell}\mathcal{N}(\Theta^{\mathcal{X}}_{ij};\Theta^{\mathcal{Y}}_{q\ell},\sigma^2_{q\ell})\Bigg] = 0 \\
&= \sum_{j\neq i}^{N}\pi_{q\ell}\left[-\delta_{ij}\tau_{iq}\tau_{j\ell}\mathcal{N}(\Theta^{\mathcal{X}}_{ij};\Theta^{\mathcal{Y}}_{q\ell},\sigma^2_{q\ell}) + \tau_{iq}\tau_{j\ell}\mathcal{N}(\Theta^{\mathcal{X}}_{ij};\Theta^{\mathcal{Y}}_{q\ell},\sigma^2_{q\ell}) + \delta_{ij}\tau_{iq}\tau_{j\ell}\right] = \sum_{j\neq i}^{N}\delta_{ij}\tau_{iq}\tau_{j\ell} \\
\hat{\pi}_{q\ell} &:= \frac{\sum_{j\neq i}^{N}\delta_{ij}\tau_{iq}\tau_{j\ell}}{\sum_{j\neq i}^{N}\left[\tau_{iq}\tau_{j\ell}\mathcal{N}(\Theta^{\mathcal{X}}_{ij};\Theta^{\mathcal{Y}}_{q\ell},\sigma^2_{q\ell}) + \delta_{ij}\tau_{iq}\tau_{j\ell} - \delta_{ij}\tau_{iq}\tau_{j\ell}\mathcal{N}(\Theta^{\mathcal{X}}_{ij};\Theta^{\mathcal{Y}}_{q\ell},\sigma^2_{q\ell})\right]}
\end{aligned}
\tag{20}
$$

□

3.3.2 Update of α:

Here our goal is to derive the update of the mixing parameter, α_q. While maintaining the variational distribution $q(A, Z)$ constant, we aim to maximize the lower bound presented in Eq.(18) concerning $\alpha_{q\ell}$. Consequently, the optimal update is obtained straightforwardly and can be expressed as:

$$\hat{\alpha}_q = \frac{1}{N}\sum_{i=1}^{N}\tau_{iq} \tag{21}$$

References

L. Li, M. Yang, C. Wang, and B. Wang. Cubature kalman filter based point set registration for slam. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pages 847–852. IEEE, 2016.

A. Myronenko and X. Song. Point set registration: Coherent point drift. *IEEE transactions on pattern analysis and machine intelligence*, 32(12):2262–2275, 2010.

Z. Zhang. Iterative point matching for registration of free-form curves and surfaces. *International journal of computer vision*, 13(2):119–152, 1994.

F. Zhou and F. De la Torre. Factorized graph matching. *IEEE transactions on pattern analysis and machine intelligence*, 38(9):1774–1789, 2015.

Z. Zhou, J. Zheng, Y. Dai, Z. Zhou, and S. Chen. Robust non-rigid point set registration using student's-t mixture model. *PloS one*, 9(3):e91381, 2014.

H. Zhu, B. Guo, K. Zou, Y. Li, K.-V. Yuen, L. Mihaylova, and H. Leung. A review of point set registration: From pairwise registration to groupwise registration. *Sensors*, 19(5):1191, 2019.

X. Zhu, M. Ding, T. Huang, X. Jin, and X. Zhang. Pcanet-based structural representation for nonrigid multimodal medical image registration. *Sensors*, 18(5):1477, 2018.