SUPPLEMENTARY MATERIALS: Probabilistic Registration for Gaussian Process 3D shape modelling in the presence of extensive missing data*

Filipa Valdeira[†], Ricardo Ferreira[‡], Alessandra Micheletti[†], and Cláudia Soares[§]

SM1. Proof for update equations in 4.2.2.

SM1.1. Proposition 1. With $p(s,t,\theta)$ as defined in (4.4) we take expectation with respect to $q_2(c,e)$ and $q_{3i}(\varsigma_i)$ and drop the terms non dependent on δ, as they are included in a constant term.

$$\mathbb{E}_{-1}[\log p(x,y,\theta)] = -\frac{1}{2}\delta^T K^{-1}\delta - \frac{1}{2}\sum_{i=1}^{N_T}\frac{1}{\varsigma_i^2}\sum_{j=1}^{N_S}\|s_j - \mathcal{T}_i\|^2 p_{ij} + \mathrm{C}.$$

Manipulating the second term, we are able to obtain

$$\begin{aligned}\sum_{i=1}^{N_T}\frac{1}{\varsigma_i^2}\sum_{j=1}^{N_S}\|s_j - \mathcal{T}_i\|^2 p_{ij} &= \sum_{i=1}^{N_T}\frac{1}{\varsigma_i^2}\|t_i+\delta_i\|^2\nu_i - 2\sum_{j=1}^{N_S}\sum_{i=1}^{N_T}\frac{1}{\varsigma_i^2}p_{ij}s_j^T(t_i+\delta_i) + C\\ &= (t+\delta)^T\tilde{D}_\nu\tilde{D}_{\varsigma^2}^{-1}(t+\delta) - 2s^T\tilde{P}^T\tilde{D}_{\varsigma^2}^{-1}(t+\delta) + C\\ &= \delta^T\tilde{D}_\nu\tilde{D}_{\varsigma^2}^{-1}\delta - 2\Big(\tilde{D}_\nu^{-1}\tilde{P}s - t\Big)^T\tilde{D}_\nu\tilde{D}_{\varsigma^2}^{-1}\delta + C.\end{aligned}$$

Therefore, the expectation becomes

$$\begin{aligned}\mathbb{E}_{-1}[\log p(x,y,\theta)] &= -\frac{1}{2}\delta^T K^{-1}\delta - \frac{1}{2}\delta^T\tilde{D}_\nu\tilde{D}_{\varsigma^2}^{-1}\delta + \Big(\tilde{D}_\nu^{-1}\tilde{P}s - t\Big)^T\tilde{D}_\nu\tilde{D}_{\varsigma^2}^{-1}\delta + C\\ &= -\frac{1}{2}\delta^T(K^{-1} + \tilde{D}_\nu\tilde{D}_{\varsigma^2}^{-1})\delta + \Big(\tilde{D}_\nu^{-1}\tilde{P}s - t\Big)^T\tilde{D}_\nu\tilde{D}_{\varsigma^2}^{-1}\delta + C\\ &= -\frac{1}{2}\delta^T\Sigma^{-1}\delta + \Big(\hat{s} - t\Big)^T\tilde{D}_\nu\tilde{D}_{\varsigma^2}^{-1}\delta + C,\end{aligned}$$

with $\Sigma = (K^{-1} + \tilde{D}_\nu\tilde{D}_{\varsigma^2}^{-1})^{-1}$ and $\hat{s} = \tilde{D}_\nu^{-1}\tilde{P}s$. Finally, we can obtain

$$\mathbb{E}_{-1}[\log p(x,y,\theta)] = -\frac{1}{2}\Big\{\delta - \Sigma\tilde{D}_\nu\tilde{D}_{\varsigma^2}^{-1}(\hat{s}-t)\Big\}^T\Sigma^{-1}\Big\{\delta - \Sigma\tilde{D}_\nu\tilde{D}_{\varsigma^2}^{-1})^{-1}(\hat{s}-t)\Big\} + \mathrm{C},$$

*Submitted to the editors DATE.

Funding: This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Project BIGMATH, Grant Agreement No 812912.

[†]Department of Environmental Science and Policy, Università degli Studi di Milano, via Celoria 2, 20133 Milan, Italy (filipa.marreiros@unimi.it,alessandra.micheletti@unimi.it).

[‡]μRoboptics, Lisbon, Portugal (ricardo.ferreira@roboptics.pt).

[§] NOVA LINCS, Computer Science Department, NOVA School of Science and Technology, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal (Claudia.soares@fct.unl.pt)

from which we have that $q_1^*(\delta)$ follows a multivariate normal distribution with the following parameters

$$q_1(\delta)^* = \phi(\delta; \Sigma\tilde{D}_\nu \tilde{D}_{\varsigma^2}^{-1}(\hat{S} - T), \Sigma),$$

meaning that

$$\begin{aligned}\mathbb{E}[\delta] &= \Sigma D_\nu D_{\varsigma^2}^{-1}(\tilde{D}_\nu^{-1}\tilde{P}s - T)\\ \mathrm{Cov}(\delta) &= \Sigma = (K^{-1} + \tilde{D}_\nu \tilde{D}_{\varsigma^2}^{-1})^{-1}.\end{aligned}$$

SM1.2. Proposition 2. We will take the expectation of the joint with respect to $q_1(\delta)$, $q_{3i}(\varsigma_i)$ and drop the terms not dependent on c and e, obtaining

$$\mathbb{E}_{-2}[\log p(x, y, \theta)] = \sum_{j=1}^{N_S} \left[\log \left(\{w p_{out}(s_j)\}^{1-c_j} \left\{ \frac{1-w}{N_T} \right\}^{c_j} \right) + \sum_{i=1}^{N_T} \log\{\langle \phi_{ij} \rangle\}^{c_j \gamma_i(e_j)} \right] + C,$$

where we defined $\langle \phi_{ij} \rangle = \exp E[\log(\phi_{ij})]$, computed from the other steps and assumed known at this point. As the the variance is only included in $\langle \phi_{ij} \rangle$, the computation of p_{ij} remains as in [SM1]

$$p_{ij} = \frac{(1-w)\langle \phi_{ij} \rangle}{N_T w p_{out}(s_j) + (1-w)\sum_{i'=1}^{N_T} \langle \phi_{i'j} \rangle}$$

and we omit the derivation. Finally, we have that

$$\begin{aligned}\langle \phi_{ij} \rangle = \exp \mathbb{E}[\log(\phi_{ij})] &= \exp\left\{ \mathbb{E}\left[\log \left(\frac{1}{(\varsigma_i\sqrt{2\pi})^D} \exp\left\{ -\frac{\|s_j - \bar{t}_i\|^2}{2\varsigma_i^2} \right\} \right) \right] \right\}\\ &= \frac{1}{(\varsigma_i\sqrt{2\pi})^D} \exp\left\{ -\frac{\|s_j - E[\bar{t}_i]\|^2 + \mathrm{Tr}(\mathrm{Cov}(\delta_i))}{2\varsigma_i^2} \right\}\\ &= \phi_{ij}(s_j; \mathbb{E}[\bar{t}_i], \varsigma_i^2) \exp\left\{ -\frac{\mathrm{Tr}(\mathrm{Cov}(\delta_i))}{2\varsigma_i^2} \right\}.\end{aligned}$$

SM1.3. Proposition 3. Because $q_{3i}(\varsigma_i)$ is a Delta dirac function we will directly maximize the lower bound, taking the expectation w.r.t. $q_1(\delta)$ and $q_2(c, e)$ and drop terms not dependent on ς_i^2. Therefore, we have

$$\begin{aligned}\mathbb{E}[\log p(x, y, \theta)] &= -\frac{D}{2}\sum_{j=1}^{N_S}\sum_{i=1}^{N_T} p_{ij}\log\varsigma_i^2 - \frac{1}{2}\sum_{i=1}^{N_T}\frac{1}{\varsigma_i^2}\sum_{j=1}^{N_S}\Big[\|s_j - E[\bar{t}_i]\|^2 + \mathrm{Tr}(\mathrm{Cov}(\delta_i))\Big]p_{ij} + C\\ &= -\frac{1}{2}\sum_{i=1}^{N_T}\left(D\nu_i \log\varsigma_i^2 + \frac{1}{\varsigma_i^2}\beta_i \right) + C\end{aligned}$$

where we have defined $\beta_i = \sum_{j=1}^{N_S} p_{ij}\big(\|s_j - \hat{t}_i\|^2 + \mathrm{Tr}(\mathrm{Cov}(t_i))\big)$, assumed known at this step. In order to maximize the ELBO we can minimize

$$\sum_{i=1}^{N_T} \Big(D\nu_i \log \varsigma_i^2 + \frac{1}{\varsigma_i^2}\beta_i\Big),$$

so we can minimize each of the terms, such that

$$(\varsigma_i^2)^* = \underset{\varsigma_i^2}{\mathrm{argmin}}\, D\log(\varsigma_i^2)\nu_i + \frac{1}{\varsigma_i^2}\beta_i.$$

Taking the derivative and equalling to zero we get

$$(\varsigma_i^2)^* = \frac{\beta_i}{D\nu_i}.$$

We can now reformulate this expression, noting that β_i can be written as

$$\begin{aligned}\beta_i &= \sum_{j=1}^{N_S} p_{ij} s_j^T s_j - 2\sum_{j=1}^{N_S} p_{ij} s_j^T \mathbb{E}[\bar{t}_i] + \sum_{j=1}^{N_S} p_{ij}\mathbb{E}[\bar{t}_i]^T\mathbb{E}[\bar{t}_i] + \sum_{j=1}^{N_S} p_{ij}\mathrm{Tr}(\mathrm{Cov}(t_i))\\ &= [P\mathrm{diag}(S)S]_i - 2\mathbb{E}[\bar{t}_i]^T[PS]_i + \nu_i\mathbb{E}[\bar{t}_i]^T\mathbb{E}[\bar{t}_i] + \nu_i\mathrm{Tr}(\mathrm{Cov}(t_i))\end{aligned}$$

and so

$$\varsigma_i^2 = \frac{1}{D}\left[\frac{[P\mathrm{diag}(S)S]_i - 2\mathbb{E}[\bar{t}_i]^T[PS]_i}{\nu_i} + \|\mathbb{E}[\bar{t}_i]\|^2 + \mathrm{Tr}(\mathrm{Cov}(\delta_i))\right].$$

SM2. Detailed settings for experimental results.

SM2.1. Experiments in Section 5.1. All methods consisting of GPSF variations (i.e. *GPSF_Full*, *GPSF_bcpdReg*, *GPSF_noTresh*) use a Squared Exponential Kernel, with a variance of 0.05 and a lengthscale of 1.5. The outlier probability is set to 0.1 and $P_{MIN} = 0.01$. The initial value for the registration variance is $\varsigma^2 = 1$. *GPClosestPnt* uses the same kernel. The variance for the observations (constant over the iterations) is set to 0.1 and the maximum distance for the closest point attribution is 0.15. The parameters for BCPD can be found in Table SM1, where we keep the notation used in the original paper and in the authors implementation. The parameters for all methods were optimized on the Fish dataset with deformation level 2, by grid search. They are kept constant throughout the experiments, except for the variation of ω when pointed out.

	ω	λ	β	γ	normalization
BCPD_Standard	0.1	2	2	3	e
BCPD_OPT_Norm	0.1	1	1.5	2	e
BCPD_OPT_noNorm	0.1	1	10	0.1	x

Table SM1
Parameters for BCPD experiments.

SM2.2. Experiments in Section 5.2.

SM2.2.1. Dataset. The Simulated dataset is obtained by applying the following deformations to the original Ear dataset. We compute the average Euclidean distance between the template and each shape in the dataset, after which we select a subset of the 15 shapes with largest distance, with an average distance of 4.35cm.

Missing data. The real scans have missing points, not only uniformly spread, but also concentrated in particular regions of the ear which are more difficult to capture by the scanning process. Therefore, in the Ear dataset we introduce both uniform and structured missing data points. The former are randomly taken from the entire point set, corresponding to 5% of the total number of points. The latter are completely removed from a predefined region.

Outliers. The ear region also contains outliers, i.e. points with no correspondence in the template. In particular, the structured outliers come from the fact that when we cut the ear portion from the entire head of the scan we do not know exactly which points belong to the ear, and consequently include some extra points. To simulate this, we define a region around the ear where outliers are added with a 0.2 ratio of the total number of points of the shape.

Measurement Noise. For each point in the Ear dataset we introduce Gaussian noise with zero mean and standard deviation of 0.07, so that they are slightly displaced, to simulate the lack of complete accuracy in the screening process.

Slight rotation, translation and scaling. Even after removing the main components of these 3 transformations it is expected that the different scans still present a small difference, not only due to limitations on the first step, but also due to natural differences in shape that do not allow for a better result. However, the Ear dataset is perfectly aligned, which can produce misleading results. Therefore, we apply to all shapes a random rotation uniformly taken from the interval of $-4°$ to $+4°$ on each axis, random scaling uniformly taken from the interval $\big[0.8, 1.2\big]$ and translation from the interval $\big[-3, 3\big]$ on each component.

SM2.2.2. Methods parameters. BCPD settings are: $\lambda = 10$, $\beta = 1$, $\omega = 0.3$, $\gamma = 0.1$ and normalization option set to x (normalized w.r.t. target shape). SFGP uses a Squared Exponential kernel with lengthscale of 10 and variance of 10. The outlier probability is set to 0.1 and $P_{MIN} = 0.01$. The initial value for the registration variance is $\varsigma^2 = 5$.

SM2.3. Implementation details.

For BCPD we use the code provided by the authors at https://github.com/ohirose/bcpd. Gaussian Process Regression in SFGP and GP Closest Point is computed with the GPFlow library [SM2].

REFERENCES

[1] O. Hirose, *A bayesian formulation of coherent point drift*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 43 (2021), pp. 2269–2286.

[2] A. G. d. G. Matthews, M. van der Wilk, T. Nickson, K. Fujii, A. Boukouvalas, P. León-Villagrá, Z. Ghahramani, and J. Hensman, *GPflow: A Gaussian process library using TensorFlow*, Journal of Machine Learning Research, 18 (2017), pp. 1–6, http://jmlr.org/papers/v18/16-537.html.