Approximating Intersections and Differences Between Linear Statistical Shape Models Using Markov Chain Monte Carlo – Supplementary Material –

In this supplementary material, we provide (i) additional information on the equivalence of the two definitions for Q in Section 3.1 of the main paper, (ii) further implementation details on our algorithm, (iii) additional information on the evaluation metrics used in Section 4.1 of the main text, (iv) an ablation study on the key parameter of our method, σ, (v) a practical strategy on how to empirically choose σ, and finally (vi) some additional qualitative results.

1. Equivalence of Q

In Section 3.1 of the main paper, the set Q was defined as

$$Q := \{x \in M \mid (x-\bar{x})^T C^{-1}(x-\bar{x}) \leq k\} \subseteq \mathbb{R}^{dn}, \quad (1)$$

where $k \in [0,1]$. We claimed that, in terms of a probabilistic interpretation, Q can be equivalently rewritten as

$$Q = \{x \in M \mid p(x) \geq \xi\} \quad (2)$$

with $\xi \in [0,1]$, see Eq. (3) of the main paper. This can be easily verified by

$$\begin{aligned}
& (x-\bar{x})^T C^{-1}(x-\bar{x}) \leq k \\
\iff \quad & -\frac{1}{2}(x-\bar{x})^T C^{-1}(x-\bar{x}) \geq -\frac{1}{2}k \\
\iff \quad & \exp\left(-\frac{1}{2}(x-\bar{x})^T C^{-1}(x-\bar{x})\right) \geq \exp\left(-\frac{1}{2}k\right) \\
\iff \quad & \frac{\exp\left(-\frac{1}{2}(x-\bar{x})^T C^{-1}(x-\bar{x})\right)}{\sqrt{(2\pi)^q \det C}} \geq \frac{\exp\left(-\frac{1}{2}k\right)}{\sqrt{(2\pi)^q \det C}} \\
\iff \quad & p(x) \geq \xi.
\end{aligned} \quad (3)$$

Hence, the hyper-ellipsoid Q contains all likely shapes, *i.e.*, those with a probability greater than a certain threshold, ξ.

2. Implementation Details

We use Markov chain Monte Carlo (MCMC) to estimate the posterior distribution $p(x \mid x \in I_\epsilon)$ for $x \in Q_1$ (and Q_2, respectively) as stated in Eq. (7) of the main paper. The Markov chain is built by means of the Metropolis-Hastings (MH) algorithm, summarized in Algorithm 1. The MH algorithms requires a proposal distribution $Q(\alpha' \mid \alpha)$, conditioned on the current state $\alpha \in \mathbb{R}^{q_1}$. We use a random walk mixture proposal of the form:

$$Q(\alpha' \mid \alpha) = \sum_{i=1}^{n} c_i Q_i(\alpha' \mid \alpha) \quad \text{with} \quad \sum_{i=1}^{n} c_i = 1. \quad (4)$$

In our specific implementation, we set $n = 4$ and defined:

$$\begin{aligned}
Q_1(\alpha' \mid \alpha) &= \mathcal{N}(\alpha, 0.2), \quad & c_1 &= 0.1, \\
Q_2(\alpha' \mid \alpha) &= \mathcal{N}(\alpha, 0.1), \quad & c_2 &= 0.5, \\
Q_3(\alpha' \mid \alpha) &= \mathcal{N}(\alpha, 0.025), \quad & c_3 &= 0.2, \\
Q_4(\alpha' \mid \alpha) &= \mathcal{N}(\|\alpha\|, 0.2), \quad & c_4 &= 0.2.
\end{aligned} \quad (5)$$

This proposal distribution was originally presented in [5] and we left it unchanged. Note that proposal is *symmetric*, *i.e.*, $Q(\alpha' \mid \alpha) = Q(\alpha \mid \alpha')$, since all mixture components are Gaussian.

Having the proposal distribution in mind, the MH algorithm proceeds as follows: In every iteration, a new sample α' is proposed based only on the previous sample, α_i.

Algorithm 1 Metropolis-Hastings algorithm (symmetric proposal distribution)

1: Initialize α_0; set $x_0 = f_1(\alpha_0)$.
2: **for** $i = 0, 1, \ldots, m$ **do**
3: Draw sample α' from $Q(\alpha' \mid \alpha_i)$; set $x' = f_1(\alpha')$.
4: Compute acceptance ratio as

$$t = \frac{L(x'; x' \in I_\epsilon)p(x')}{L(x_i; x_i \in I_\epsilon)p(x_i)}.$$

5: Accept α' with probability t by drawing a sample r from $\mathcal{U}(0,1)$ and

$$\alpha_{i+1} = \begin{cases} \alpha' & \text{if } t > r, \\ \alpha_i & \text{otherwise.} \end{cases}$$

6: **end for**
7: **return** $\{f_1(\alpha_0), f_1(\alpha_1), \ldots, f_1(\alpha_m)\}$

The proposed sample is then either accepted or rejected with probability t, where t is the so-called acceptance ratio. After a sufficient number of iterations, m, the MH algorithm returns a set of *accepted* samples from the desired posterior distribution, $p(x \mid x \in I_\epsilon)$.

3. Evaluation Metrics

We now provide some additional information on the evaluation metrics used for quantitative analysis as briefly explained in Section 4.1 of the main paper.

Grassmann distance. The Grassmann distance is the natural distance between two *linear* subspaces embedded in $\mathbb{R}^n$ (the set of all k-dimensional linear subspaces is called the Grassmannian, usually denoted as $\mathrm{Gr}(k, n)$). Given orthonormal bases $A, B \in \mathbb{R}^{k \times n}$ for two subspaces from $\mathrm{Gr}(k, n)$, the Grassmann distance can be calculated by means of the principal angles $\{\theta_1, \theta_2, \ldots, \theta_k\}$ between the two subspaces. With slight abuse of notation but for the sake of brevity, we refer to the Grassmann distance as $d_G(A, B) = \|(\theta_1, \theta_2, \ldots, \theta_k)\|_2$.

Since SSMs span *affine* subspaces, the Grassmann distance as presented previously can not directly be applied to measure distances between subspaces spanned by linear shape models. Fortunately, as shown in [4] (Theorem 7), the Grassmann distance can be easily extended to affine subspaces as we briefly explain next. The key idea is to embed the affine subspace into a linear subspace by adding one dimension. Given two affine subspaces represented by orthonormal bases $A, B \in \mathbb{R}^{k \times n}$ and displacement vectors $b, c \in \mathbb{R}^n$, their Stiefel coordinates, $Y_1, Y_2 \in \mathbb{R}^{(k+1) \times n}$ are given by

$$Y_1 = \begin{pmatrix} A & b_0 / \sqrt{1 + \|b_0\|^2} \\ 0 & 1 / \sqrt{1 + \|b_0\|^2} \end{pmatrix} \tag{6}$$

and

$$Y_2 = \begin{pmatrix} B & c_0 / \sqrt{1 + \|c_0\|^2} \\ 0 & 1 / \sqrt{1 + \|c_0\|^2} \end{pmatrix}, \tag{7}$$

where b_0 and c_0 are unit vectors orthogonal to the columns of A and B, respectively. We compute them by

$$b' = b - \sum_{j=1}^{k} (a_j \cdot b) a_j, \quad b_0 = \frac{b'}{\|b'\|_2}, \tag{8}$$

and analogously for c_0. Here, $a_j \in \mathbb{R}^n$, $j = 1, \ldots, k$, denote the columns of A. Finally, the affine Grassmann distance is computed by applying Singular Value Decomposition to $Y_1^T Y_2$, yielding the $k + 1$ principal angles between the respective affine subspaces. Taking the Euclidean norm of those angles leads to the affine Grassmann distance.

Reconstruction error. Following the main paper, to evaluate the quality of the computed differences, we exploit the fact that we have samples $\{x_1, x_2, \ldots, x_r\}$ from the ground-truth difference, D_{12} (or D_{21}). Denote the

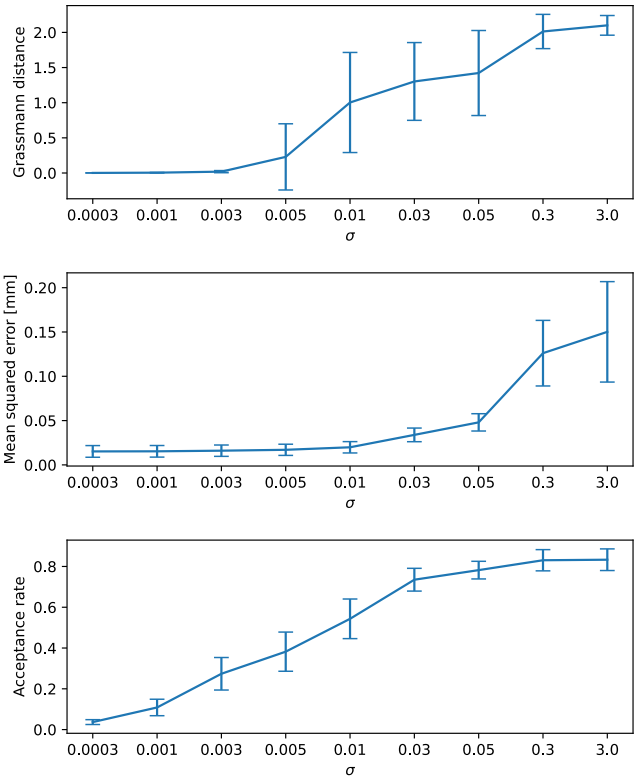

Figure 1. Results of the ablation study for different values of σ, averaged over all six star models (described in the main paper). Top: Grassmann distance between estimated intersection space and ground-truth intersection; center: mean squared error between MCMC samples and their corresponding projections into the other model (averaged over all posterior samples); bottom: acceptance rates during MCMC sampling.

MCMC samples from the estimated difference $\hat{D}_{12}$ as $\{\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_r\}$. We then evaluate the quality of $\hat{D}_{12}$ by quantifying whether or not samples $\hat{x}_j$ can be as badly represented in the ground-truth intersection, I, as samples x_j from the ground-truth difference. Here, we expect high errors since shapes belong to the difference if they can *not* be represented in the intersection. Formally, we calculate the reconstruction errors

$$d_R(D_{12}, I) = \frac{1}{r} \sum_{j=1}^{r} d(\mathrm{proj}_{M_I}(x_i), x_i) \tag{9}$$

and

$$d_R(\hat{D}_{12}, I) = \frac{1}{r} \sum_{j=1}^{r} d(\mathrm{proj}_{M_I}(\hat{x}_i), \hat{x}_i) \tag{10}$$

by projecting samples onto the subspace M_I spanned by I and evaluating its distance using the mean squared error (MSE; see Eq. (10) of the main paper).

Figure 2. Random samples from the male and female color model of the LYHM [1] data set (1st and 2nd block) as well as samples of the computed intersection space (3rd block), and respective differences (male without female, and female without male, 4th and 5th block). Similar to shape (see Figure 3 of the main paper), we see stronger male and female dominance in the differences and neutral gender in the intersection, especially visible in the beard region. All color samples are visualized on the mean face.

4. Ablation Study

We identified the variance involved in the distance likelihood, σ^2, as the most important parameter of our algorithm (see Eq. (9) of the main paper). To study its effects, in addition to the quantitative evaluation presented in the main text, we also provide an ablation study for different values of σ. The ablation is carried out on the star data set as described in Section 4.1.1 of the main paper (see also Table 1 of the main paper). Moreover, we only investigate the estimation of ground-truth intersection spaces as our method shows similar behavior for differences.

The results can be found in Figure 1 (top row), averaged over all six star models. As seen, setting σ too high leads to large Grassmann distances, implying a worse estimation of the ground-truth intersection. Contrary, the smaller σ, the better the estimation of the true intersection space. Interestingly, starting from $\sigma = 0.003$, its exact value seems to become less critical as even a decrease of factor 10 does not lead to significant changes.

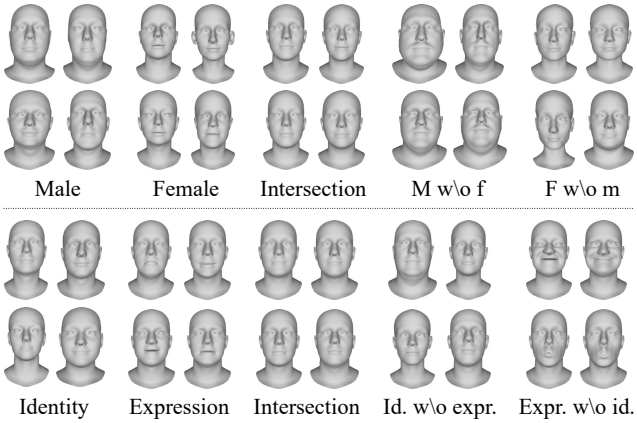

Figure 3. Random samples from the male and female models (top) and identity and expression models (bottom) of FLAME (1st and 2nd block). Also shown are samples of the computed intersection model (3rd block) as well as from the difference between male and female (top) and identity and expression models (bottom; 4th and 5th block). We observe very similar results as for LYHM models and BFM 2019 [2], see Figures 3 and 4 of the main paper.

5. How to Choose σ In Practice?

Since we usually do not have ground-truth intersections for real-world SSMs, it is natural to wonder how to empirically validate the performance of our method and the chosen σ in practice. To this end, we also report the MSE between an MCMC sample, $x \in Q_1$ (or $x \in Q_2$), and its projection into the other model, x', as well as the acceptance rates during MCMC sampling, see Figure 1 (center and bottom row). Note that the MSE is computed between *all* posterior samples and their respective projections; the average thus serves as an empirical estimation for the *mean* ϵ in Eqs. (5) and (6) of the main text.

In terms of MSE, we observe an almost similar behavior as for the Grassmann distances. Smaller MSEs correspond to lower Grassmann distances. As a result of this correlation, the distance between x and x' can be used as an indication to determine a suitable value for σ. It does not require a ground-truth and can be easily monitored during the run time of MCMC. To avoid setting σ too small (and preventing incorrect estimation of the desired distribution), however, one should ensure that the acceptance rates are between 0.25 and 0.5, see [5].

In conclusion, although a suitable value for σ might not be trivial to determine in practice, it can be well chosen by carefully inspecting the MSE between posterior samples and their corresponding projections as well as acceptance rates during MCMC sampling. Both quantities can be easily computed without the necessity of ground-truth intersections or differences.

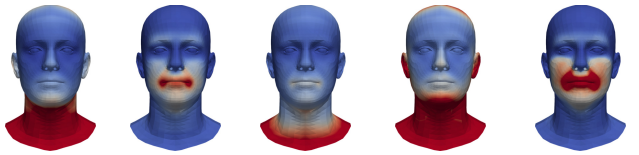

Figure 4. Per-vertex level visualization of similarities and differences between the identity and expression spaces of FLAME. Depicted is the per-vertex variance of identity and expression models as well as the posterior variance for the computed intersection model and differences (red high variance, blue low). We observe a similar behavior as for BFM 2019, see Figure 5 of the main paper.

6. Additional Qualitative Results

In this section, we provide results for the extension to color experiment (Section 4.3 of the main paper) as well as additional qualitative results on the FLAME [3] model.

Extension to color. The results are shown in Figure 2. We observe very similar behavior for the color as we have seen for shape, see Figure 3 of the main paper. Male features are exaggerated in faces drawn from the difference of male and female, and samples from the difference of female and male appear more feminine than faces from the original female model. Similar to the effect in shape, we also perceive neutral textures in the intersection space.

Results on FLAME. We show exemplary results on FLAME in Figures 3 and 4. They are very similar to the findings presented in the main paper, please refer to Figures 3–5 of the main paper.

References

[1] Hang Dai, Nick Pears, William AP Smith, and Christian Duncan. Statistical modeling of craniofacial shape and texture. *IJCV*, 28(2):547–571, 2019. 3

[2] Thomas Gerig, Andreas Morel-Forster, Clemens Blumer, Bernhard Egger, Marcel Lüthi, Sandro Schönborn, and Thomas Vetter. Morphable face models-an open framework. In *FG*, pages 75–82, 2018. 3

[3] Tianye Li, Timo Bolkart, Michael J Black, Hao Li, and Javier Romero. Learning a model of facial shape and expression from 4d scans. In *ACM TOG*, pages 194:1–194:17, 2017. 4

[4] Lek-Heng Lim, Ken Sze-Wai Wong, and Ke Ye. The grassmannian of affine subspace. *Found. Comput. Math.*, 21(2):537–574, 2020. 2

[5] Sandro Schönborn, Bernhard Egger, Andreas Morel-Forster, and Thomas Vetter. Markov chain monte carlo for automated face image analysis. *IJCV*, 123(2):160–183, 2017. 1, 3