

Supplementary Material for A Groupwise Multilinear Correspondence Optimization for 3D Faces

Timo Bolkart
Saarland University, Germany
tbolkart@mnci.uni-saarland.de

Stefanie Wuhrer
Inria Rhône-Alpes, France
stefanie.wuhrer@inria.fr

Given a set of multilinearly distributed data, our submission proposes an approach that produces a high-quality registration along with an improved multilinear model. This appendix provides more details on how the objective function that measures model quality depending on correspondences is optimized in Section 1. Further, the computational complexities of the multilinear method and existing linear approaches are compared in Section 2.

1. Formulation of optimization

This section derives the objective function used in our submission. Let $\mathbf{x}_{ie} \in \mathbb{R}^{3n}$ denote the face of identity i in expression e that consists of n vertices. Let $\mathbf{v}_k(\mathbf{x}_{ie}) \in \mathbb{R}^3$ denote the k -th vertex of \mathbf{x}_{ie} . To simplify notation, whenever a fixed shape is used, we omit the subscripts ie . Each face is centered by subtracting the mean face $\bar{\mathbf{x}} = \frac{1}{d_2 d_3} \sum_{i=1}^{d_2} \sum_{e=1}^{d_3} \mathbf{x}_{ie}$. Let $\mathbf{c}_{ie} := \mathbf{x}_{ie} - \bar{\mathbf{x}}$ denote the centered face.

Recall that we minimize the groupwise objective function

$$E = E_{COMP} + w_{REG} E_{REG}, \quad (1)$$

where E_{COMP} measures the compactness, the regularizer E_{REG} encourages the points to be regularly distributed over the mesh, and w_{REG} is a weight that controls the influence of the regularization.

As described in the submission, the compactness energy is

$$E_{COMP} = \frac{1}{d_2} \sum_{a=1}^{d_2} \underbrace{\ln(\lambda_a^{(2)} + \delta_2)}_{E_{COMP,a}^{(2)}} + \frac{1}{d_3} \sum_{a=1}^{d_3} \underbrace{\ln(\lambda_a^{(3)} + \delta_3)}_{E_{COMP,a}^{(3)}}, \quad (2)$$

where $\lambda_a^{(n)}$ denotes the a -th eigenvalue of the mode- n covariance matrix. Small regularization constants δ_n are used to avoid singularities of E_{COMP} for vanishing eigenvalues. Equivalent to the higher order singular value decomposition (HOSVD) [6], the mode-2 and mode-3 covariance matrices are computed as $\mathbf{D}^{(2)} = \frac{1}{d_3} \mathbf{A}_{(2)} \mathbf{A}_{(2)}^T$ and $\mathbf{D}^{(3)} = \frac{1}{d_2} \mathbf{A}_{(3)} \mathbf{A}_{(3)}^T$, respectively. Here, $\mathbf{A}_{(n)}$ denotes the matrix unfolding of the tensor \mathcal{A} in the direction of n -th mode, where \mathcal{A} contains the centered data \mathbf{c}_{ie} . The regularization energy for each shape is

$$E_{REG} = \frac{1}{n} \sum_{k=1}^n \|U^2(\mathbf{v}_k(\mathbf{x}))\|^2. \quad (3)$$

The double-umbrella operator $U^2(\mathbf{p})$ is the discrete bi-Laplacian approximation [4] computed by

$$U^2(\mathbf{p}) = \frac{1}{|N(\mathbf{p})|} \sum_{\mathbf{p}_r \in N(\mathbf{p})} U(\mathbf{p}_r) - U(\mathbf{p}), \quad (4)$$

where $N(\mathbf{p})$ denotes the set of neighbors of vertex \mathbf{p} within the mesh, and $U(\mathbf{p}) = 1/|N(\mathbf{p})| \sum_{\mathbf{p}_r \in N(\mathbf{p})} \mathbf{p}_r - \mathbf{p}$.

The objective function E in Equation 1 is analytically differentiable with respect to the 2D shape parameters α . For a fixed shape \mathbf{x} , the gradient $\frac{\partial E}{\partial \alpha_k} \in \mathbb{R}^2$ of E with respect to the parameters α_k of vertex $\mathbf{v}_k(\mathbf{x})$ is

$$\frac{\partial E}{\partial \alpha_k} = \frac{\partial \mathbf{x}}{\partial \alpha_k} \frac{\partial E_{COMP}}{\partial \mathbf{x}} + w_{REG} \frac{\partial \mathbf{x}}{\partial \alpha_k} \frac{\partial E_{REG}}{\partial \mathbf{x}}. \quad (5)$$

In the following, we provide derivations for these partial derivatives. Section 1.1 derives $\frac{\partial E_{REG}}{\partial \mathbf{x}}$ and gives the result in Equation 8, Section 1.2 derives $\frac{\partial E_{COMP}}{\partial \mathbf{x}}$ and gives the result in Equation 9, and Section 1.3 derives $\frac{\partial \mathbf{x}}{\partial \alpha_k}$ and gives the result in Equation 21.

1.1. Regularization derivative

To compute the derivative $\frac{\partial E_{REG}}{\partial \mathbf{x}}$, we first define a block matrix $\mathbf{S} \in \mathbb{R}^{3n \times 3n}$. One submatrix $\mathbf{S}_{jk} \in \mathbb{R}^{3 \times 3}$ of \mathbf{S} is defined as

$$\mathbf{S}_{jk} = \begin{cases} -\mathbf{I} & j = k \\ \frac{1}{|N(\mathbf{p}_j)|} \mathbf{I} & \mathbf{p}_k \in N(\mathbf{p}_j) \\ \mathbf{0} & otherwise, \end{cases} \quad (6)$$

where $\mathbf{I} \in \mathbb{R}^{3 \times 3}$ denotes the identity matrix.

The energy E_{REG} can then be expressed as

$$E_{REG} = \frac{1}{n} (\mathbf{S}\mathbf{x})^T (\mathbf{S}\mathbf{x}). \quad (7)$$

The derivative $\frac{\partial E_{REG}}{\partial \mathbf{x}}$ is then

$$\frac{\partial E_{REG}}{\partial \mathbf{x}} = \frac{2}{n} (\mathbf{S}\mathbf{S})^T (\mathbf{S}\mathbf{x}). \quad (8)$$

1.2. Compactness derivative

This section derives $\frac{\partial E_{COMP}}{\partial \mathbf{x}}$. For PCA models, the compactness energy is measured by the trace of the covariance matrix. Instead of minimizing the trace of the covariance matrix, Kotcheff and Taylor [5] minimize the logarithm of the determinant of the covariance matrix. They show that minimizing the logarithm of the determinant of the covariance matrix leads to a better model than minimizing the trace of the covariance matrix. This observation can be explained because the resulting energy is a log-sum penalty function on the eigenvalues of the covariance matrix, which is known to encourage sparsity [1] as small eigenvalues are heavily punished. Inspired by this method, our submission derives a compactness energy for multilinear models.

Due to the different centering, the extension of the PCA compactness gradient to the multilinear case is not straightforward. The derivative of E_{COMP} (Equation 2) with respect to the shape \mathbf{x} is by the chain rule

$$\frac{\partial E_{COMP}}{\partial \mathbf{x}} = \frac{1}{d_2} \sum_{a=1}^{d_2} \underbrace{\frac{\partial \lambda_a^{(2)}}{\partial \mathbf{x}}}_{Eq. 14} \underbrace{\frac{\partial E_{COMP,a}^{(2)}}{\partial \lambda_a^{(2)}}}_{Eq. 10} + \frac{1}{d_3} \sum_{a=1}^{d_3} \underbrace{\frac{\partial \lambda_a^{(3)}}{\partial \mathbf{x}}}_{Eq. 17} \underbrace{\frac{\partial E_{COMP,a}^{(3)}}{\partial \lambda_a^{(3)}}}_{Eq. 16}, \quad (9)$$

with different partial derivatives for the identity mode and the expression mode.

Mode-2: The derivative for mode-2 is

$$\frac{\partial E_{COMP,a}^{(2)}}{\partial \lambda_a^{(2)}} = \frac{1}{\lambda_a^{(2)} + \delta_2}, \quad (10)$$

and by the chain rule

$$\frac{\partial \lambda_a^{(2)}}{\partial \mathbf{x}} = \sum_{j=1}^{d_2} \sum_{k=1}^{d_2} \frac{\partial \mathbf{D}^{(2)}[j, k]}{\partial \mathbf{x}} \frac{\partial \lambda_a^{(2)}}{\partial \mathbf{D}^{(2)}[j, k]}, \quad (11)$$

where $\mathbf{D}^{(2)}[j, k] \in \mathbb{R}$ denotes the element of row j and column k of the mode-2 covariance matrix $\mathbf{D}^{(2)}$.

It follows from infinitesimal considerations [2, Appendix B.2] that the partial derivative of the eigenvalue $\lambda_a^{(2)}$ with respect to the element of the covariance matrix $\mathbf{D}^{(2)}[j, k]$ is

$$\frac{\partial \lambda_a^{(2)}}{\partial \mathbf{D}^{(2)}[j, k]} = \mathbf{e}_a^{(2)}[j] \mathbf{e}_a^{(2)}[k], \quad (12)$$

where $\mathbf{e}_a^{(2)}[j]$ denotes the j -th element, and $\mathbf{e}_a^{(2)}[k]$ the k -th element of the corresponding eigenvector of $\lambda_a^{(2)}$.

The partial derivative of the element of the covariance matrix $\mathbf{D}^{(2)}[j, k]$ with respect to the shape \mathbf{x} is computed with the chain rule as

$$\frac{\partial \mathbf{D}^{(2)}[j, k]}{\partial \mathbf{x}} = \sum_{l=1}^{d_2} \sum_{m=1}^{d_3} \frac{\partial \mathbf{c}_{lm}}{\partial \mathbf{x}} \frac{\partial \mathbf{D}^{(2)}[j, k]}{\partial \mathbf{c}_{lm}}. \quad (13)$$

This leads to the overall partial derivative of the eigenvalue $\lambda_a^{(2)}$ with respect to the shape \mathbf{x}_{ie} . That is,

$$\frac{\partial \lambda_a^{(2)}}{\partial \mathbf{x}_{ie}} = \frac{2}{d_2 d_3 d_3} \sum_{j=1}^{d_2} \left(\mathbf{e}_a^{(2)}[j] \sum_{k=1}^{d_2} \left(\mathbf{e}_a^{(2)}[k] \sum_{m=1}^{d_3} \Psi_{ik}^{em} \mathbf{c}_{jm} \right) \right), \quad (14)$$

where

$$\Psi_{ik}^{em} = \begin{cases} d_2 d_3 - 1 & i = k \text{ and } e = m \\ -1 & \text{otherwise,} \end{cases} \quad (15)$$

and where $\mathbf{e}_a^{(2)}[j]$ denotes the j -th element, and $\mathbf{e}_a^{(2)}[k]$ the k -th element of the corresponding eigenvector of $\lambda_a^{(2)}$.

Mode-3: The derivative for mode-3 is

$$\frac{\partial E_{COMP,a}^{(3)}}{\partial \lambda_a^{(3)}} = \frac{1}{\lambda_a^{(3)} + \delta_3}, \quad (16)$$

and using a similar argument as above,

$$\frac{\partial \lambda_a^{(3)}}{\partial \mathbf{x}_{ie}} = \frac{2}{d_2 d_3 d_2} \sum_{j=1}^{d_3} \left(\mathbf{e}_a^{(3)}[j] \sum_{k=1}^{d_3} \left(\mathbf{e}_a^{(3)}[k] \sum_{m=1}^{d_2} \Psi_{im}^{ek} \mathbf{c}_{mj} \right) \right), \quad (17)$$

where

$$\Psi_{im}^{ek} = \begin{cases} d_2 d_3 - 1 & i = m \text{ and } e = k \\ -1 & \text{otherwise,} \end{cases} \quad (18)$$

and where $\mathbf{e}_a^{(3)}[j]$ denotes the j -th element, and $\mathbf{e}_a^{(3)}[k]$ the k -th element of the corresponding eigenvector of $\lambda_a^{(3)}$.

1.3. Parametrization derivative

This section derives $\frac{\partial \mathbf{x}}{\partial \alpha_k}$. Recall that any kind of continuous mapping can be used to parametrize the shape \mathbf{x} in 2D. We establish a continuous mapping from a 3D face to a 2D unit square by a thin-plate spline [3]. For other mappings, the term $\frac{\partial \mathbf{x}}{\partial \alpha} \in \mathbb{R}^{2n \times 3n}$ of the derivative changes, while the rest of the gradient stays unchanged.

A thin-plate spline is computed for each shape as

$$\Phi(\alpha) = \mathbf{c} + \mathbf{A}\alpha + \mathbf{W}^T (\sigma(\alpha - \mathbf{t}_1), \dots, \sigma(\alpha - \mathbf{t}_n))^T, \quad (19)$$

where $\mathbf{c} \in \mathbb{R}^3$, $\mathbf{A} \in \mathbb{R}^{3 \times 2}$, and $\mathbf{W} \in \mathbb{R}^{n \times 3}$ are the parameters of the mapping, and where $\sigma : \mathbb{R}^2 \rightarrow \mathbb{R}$ is the function

$$\sigma(\mathbf{h}) = \begin{cases} \|\mathbf{h}\|^2 \log(\|\mathbf{h}\|) & \|\mathbf{h}\| > 0, \\ 0 & \|\mathbf{h}\| = 0, \end{cases} \quad (20)$$

where $\|\mathbf{h}\|$ is the Euclidean length of \mathbf{h} .

To find the derivative $\frac{\partial \mathbf{x}}{\partial \alpha_k}$, we can compute $\frac{\partial \mathbf{v}_b(\mathbf{x})}{\partial \alpha_k}$ for every vertex $\mathbf{v}_b(\mathbf{x})$ and combine the resulting derivatives. It is

$$\frac{\partial \mathbf{v}_b(\mathbf{x})}{\partial \alpha_k} = \frac{\partial \Phi(\alpha_b)}{\partial \alpha_k} = \begin{cases} \mathbf{A}^T + \left(\frac{\partial \sigma(\alpha_b - \mathbf{t}_1)}{\partial \alpha_k} \dots \frac{\partial \sigma(\alpha_b - \mathbf{t}_n)}{\partial \alpha_k} \right) \mathbf{W} & x_k = x_b \\ \mathbf{0} & x_k \neq x_b \end{cases} \quad (21)$$

with

$$\frac{\partial \sigma(\alpha - \mathbf{t}_l)}{\partial \alpha} = \begin{cases} (\alpha - \mathbf{t}_l)(2 \log \|\alpha - \mathbf{t}_l\| + 1) & \|\alpha - \mathbf{t}_l\| > 0 \\ \mathbf{0} & \|\alpha - \mathbf{t}_l\| = 0. \end{cases} \quad (22)$$

2. Comparison of computational complexities

This section gives the computational complexities for the correspondence optimization for the multilinear case and the linear case ($d_3 = 1$). For both cases, we use that computing the singular values and singular vectors of a $m \times n$ matrix takes $O(mn^2 + n^3)$ time [7, Chapter 31].

2.1. Multilinear optimization

Derivative: The partial derivatives of the eigenvalues $\lambda_a^{(2)}$ and $\lambda_a^{(3)}$ with respect to the shape \mathbf{x}_{ie} (Equations 14 and 17) can be reformulated as

$$\frac{\partial \lambda_a^{(2)}}{\partial \mathbf{x}_{ie}} = \underbrace{\frac{2}{d_3} \mathbf{e}_a^{(2)}[i] \sum_{j=1}^{d_2} \mathbf{e}_a^{(2)}[j] \mathbf{c}_{je}}_{O(nd_2)} - \frac{2}{d_2 d_3 d_3} \underbrace{\left(\sum_{k=1}^{d_2} \mathbf{e}_a^{(2)}[k] \right)}_{O(d_2)} \underbrace{\left(\sum_{j=1}^{d_2} \sum_{m=1}^{d_3} \mathbf{e}_a^{(2)}[j] \mathbf{c}_{jm} \right)}_{O(nd_2 d_3)}, \quad (23)$$

and

$$\frac{\partial \lambda_a^{(3)}}{\partial \mathbf{x}_{ie}} = \underbrace{\frac{2}{d_2} \mathbf{e}_a^{(3)}[e] \sum_{j=1}^{d_3} \mathbf{e}_a^{(3)}[j] \mathbf{c}_{ij}}_{O(nd_3)} - \frac{2}{d_2 d_3 d_2} \underbrace{\left(\sum_{k=1}^{d_3} \mathbf{e}_a^{(3)}[k] \right)}_{O(d_3)} \underbrace{\left(\sum_{j=1}^{d_3} \sum_{m=1}^{d_2} \mathbf{e}_a^{(3)}[j] \mathbf{c}_{mj} \right)}_{O(nd_2 d_3)}. \quad (24)$$

Computational complexity: The gradient E_{COMP} with respect to the shape \mathbf{x} (Equation 9) can therefore be computed in time $O(nd_2^2 d_3 + nd_2 d_3^2)$. Computing the eigenvalues and eigenvectors of the mode-2 and mode-3 covariance matrices $\mathbf{D}^{(2)}$ and $\mathbf{D}^{(3)}$ takes $O(nd_2^2 d_3 + d_3^3)$ and $O(nd_2 d_3^2 + d_3^3)$ time, respectively. This leads to the overall computational complexity of $O(nd_2^2 d_3 + nd_2 d_3^2 + d_3^3 + d_3^3)$.

Assuming $n \gg d_2, d_3$ the complexity becomes $O(nd_2^2 d_3 + nd_2 d_3^2)$.

2.2. Linear optimization

Derivative: Most previous methods use a linear model for correspondence optimization (see *e.g.* [2, Chapter 4]). Note that the linear model is a special case of our multilinear approach, where the multilinear model degenerates to the linear model for $d_2 = 1$ or $d_3 = 1$. We use a consistent notation to previous sections but omit the subscripts and superscripts for mode 2 and mode 3, since only one mode is present in the linear case.

Let d denote the number of shapes, where $\mathbf{x}_j \in \mathbb{R}^{3n}$ denotes the j -th shape, and $\mathbf{c}_j \in \mathbb{R}^{3n}$ denotes the j -th centered shape. The derivative of E_{COMP} with respect to the shape \mathbf{x} is

$$\frac{\partial E_{COMP}}{\partial \mathbf{x}} = \frac{1}{d} \sum_{a=1}^d \frac{\partial \lambda_a}{\partial \mathbf{x}} \frac{\partial E_{COMP,a}}{\partial \lambda_a}. \quad (25)$$

From the centering of the data ($\sum_{j=1}^d \mathbf{c}_j = 0$) follows

$$\sum_{j=1}^d \mathbf{e}_a[j] = 0. \quad (26)$$

The partial derivative of the eigenvalue λ_a with respect to the shape \mathbf{x}_j is therefore

$$\frac{\partial \lambda_a}{\partial \mathbf{x}_j} = 2 \mathbf{e}_a[j] \sum_{k=1}^d \mathbf{e}_a[k] \mathbf{c}_k. \quad (27)$$

Computational complexity: The partial derivative of λ_a with respect to \mathbf{x}_j (Equation 27) can be computed in time $O(nd)$. The gradient of E_{COMP} with respect to \mathbf{x} (Equation 25) can therefore be computed in time $O(nd^2)$. Computing the eigenvalues and eigenvectors of the covariance matrix \mathbf{D} takes $O(nd^2 + d^3)$ time. This leads to the overall computational complexity of $O(nd^2 + d^3)$.

Assuming $n \gg d$ the complexity becomes $O(nd^2)$.

2.3. Comparison

For both existing linear methods and our method, the minimum description length optimization is non-linear and solved with the help of optimizers that require an explicit gradient computation in each iteration. Hence, the computational complexity of the gradient computation has a strong influence on the overall run time of the optimization.

For the same number of shapes $d = d_2 d_3$, it takes time $O(nd_2^2 d_3^2)$ to compute a gradient for existing linear methods. Our multilinear model, where a gradient computation takes time $O(nd_2^2 d_3 + nd_2 d_3^2)$, is significantly more efficient if both d_2 and d_3 are large.

References

- [1] E. Candès, M. Wakin, and S. Boyd. Enhancing sparsity by reweighted l_1 minimization. *JFAA*, 14(5-6):877–905, 2008.
- [2] R. Davies, C. Twining, and C. Taylor. *Statistical Models of Shape: Optimisation and Evaluation*. Springer, 2008.
- [3] I. Dryden and K. Mardia. *Statistical shape analysis*. Wiley, 1998.
- [4] L. Kobbelt, S. Campagna, J. Vorsatz, and H.-P. Seidel. Interactive multi-resolution modeling on arbitrary meshes. In *SIGGRAPH*, pages 105–114, 1998.
- [5] A. C. Kotcheff and C. J. Taylor. Automatic construction of eigenshape models by direct optimization. *Med. Image Anal.*, 2(4):303 – 314, 1998.
- [6] L. D. Lathauwer. *Signal processing based on multilinear algebra*. PhD thesis, K.U. Leuven, Belgium, 1997.
- [7] L. N. Trefethen and D. Bau. *Numerical Linear Algebra*. SIAM, 1997.