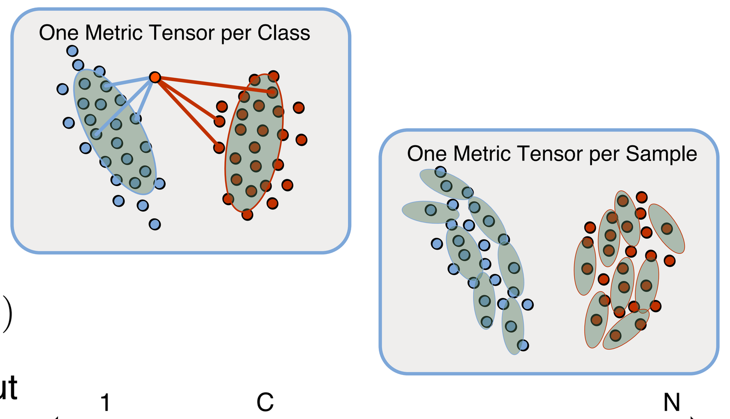
A Geometric Take on Metric Learning

Søren Hauberg[1], Oren Freifeld[2] and Michael J. Black[1]

[1]Perceiving Systems, Max Planck Institute for Intelligent Systems, Tübingen, Germany
[2]Applied Mathematics, Brown University, Providence, US

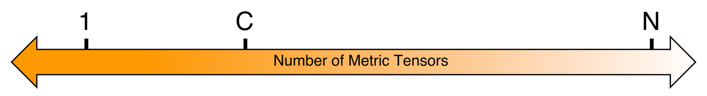
1 Multi-metric Learning

Multi-metric learning uses a set of metric tensors such that different distance measures are applied in different regions of the feature space. These are used with kNN classifiers where

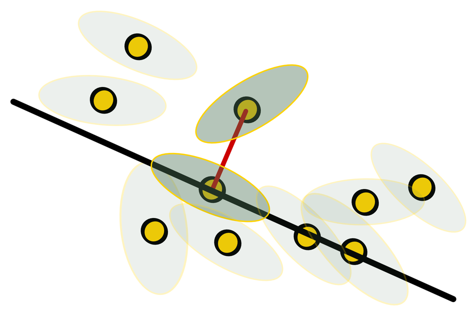


$$\mathrm{dist}^2(\mathbf{a},\mathbf{b}) = (\mathbf{a}-\mathbf{b})^T\mathbf{M}_\mathbf{a}(\mathbf{a}-\mathbf{b})$$

This has proven to be effective, but has only been applicable to classification with kNN classifiers.



We extend the idea to work with PCA and regression!



What is the problem?

The above approach

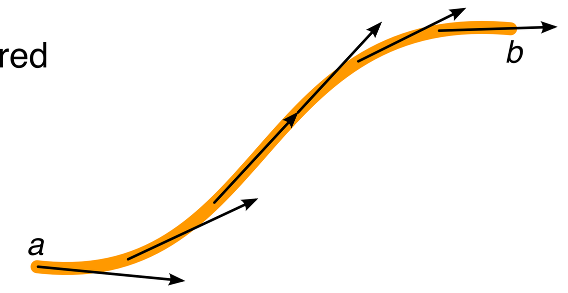
- is not metric;
- is asymmetric;
- violates the triangle inequality;
- projections are hard to define.

What is the length of the red line? And should it even be a line? All of this is confusing as we haven't defined the metric in the entire feature space.

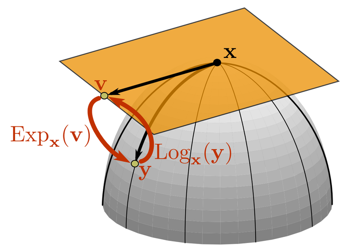
3 Riemannian Statistics

In Riemannian spaces curve length is measured by integrating the curve derivative measured according to the local metric tensor

$$\mathrm{Length}(c) = \int_0^1 \|c'(t)\|_{\mathbf{M}(c(t))}\mathrm{d}t$$



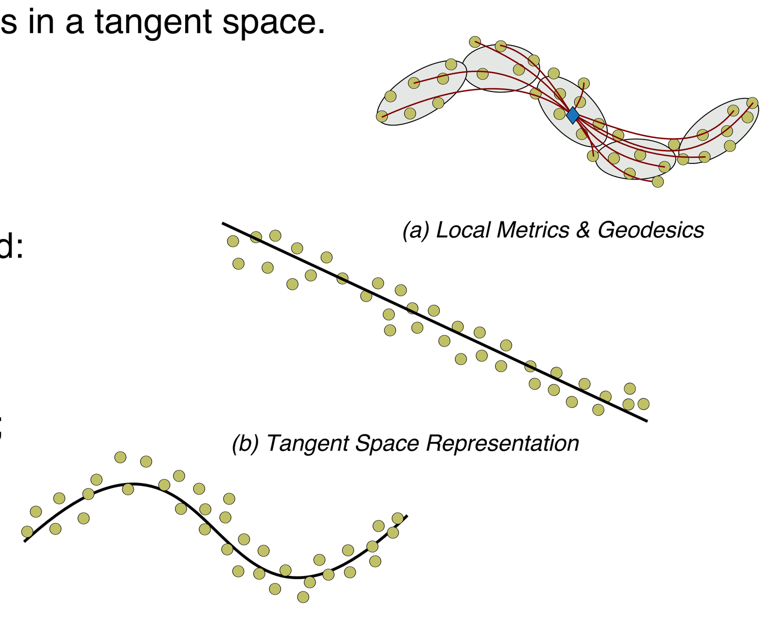
Geodesics are found as length-minimizing curves.



Mapping exists back and forth between the local tangent space of the manifold. These allow for statistics on manifolds by performing Euclidean statistics in a tangent space.

The figures to the right show the basic steps of PCA on a manifold:

- Compute the mean;
- Map data to tangent space and learn Euclidean model;
- Map Euclidean model back to the feature space.



(a) Local Metrics & Geodesics

(b) Tangent Space Representation

(c) First Principal Geodesic

2 A Metric Feature Space

A big problem with the above approach is that the metric tensor is not directly defined at all points in the image. We define it in the obvious way

$$\mathbf{M}(x) = \sum_{r=1}^{R} \frac{\tilde{w}_r(x)}{\sum_j \tilde{w}_j(x)}\mathbf{M}_r$$

where the weights are squared exponentials

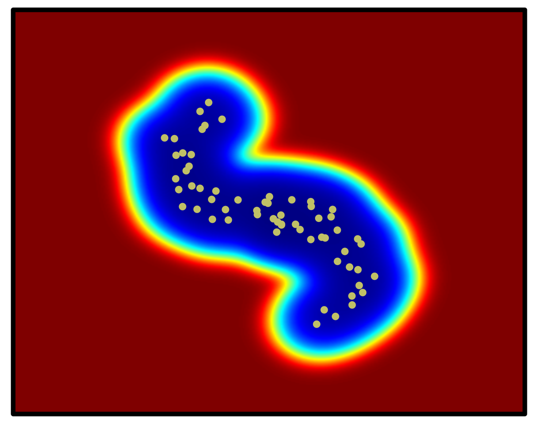
$$\tilde{w}_r(x) = \exp\left(-\frac{\rho}{2}\|x - x_r\|^2_{\mathbf{M}_r}\right)$$

With this choice, the metric tensor changes smoothly, which gives us the following result

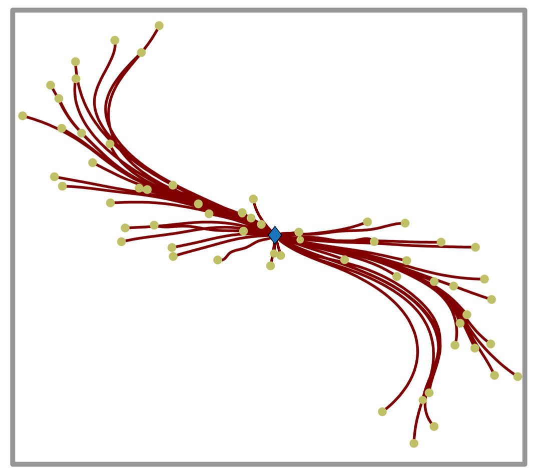
Lemma 1:
The space $\mathbb{R}^D$ endowed with the metric tensor $\mathbf{M}(x)$ is a chart of a Riemannian manifold.

This simple observation provides both theoretical insights and practical tools

- Riemannian manifolds are metric spaces, so our feature space is also metric.
- Riemannian manifolds have well-defined notions of statistics (see box 3).
- Geodesics and relevant maps are computable in practice (see next).



Example of a smooth metric; here we show the trace of the metric tensor at every point. Blue indicates 'cheap' distances while red is 'expensive'.



Example geodesic paths according to the metric tensor from the above example.

As the space is smooth we can use the Euler–Lagrange equation to compute shortest paths (geodesics)

Theorem 2:
Geodesics according to the metric satisfy the following system of 2nd order ODE's

$$\mathbf{M}(c(t))c''(t) = -\frac{1}{2}\left[\frac{\partial\mathrm{vec}\left[\mathbf{M}(c(t))\right]}{\partial c(t)}\right]^T (c'(t)\otimes c'(t))$$

The above result is general; for our squared exponential weights we get

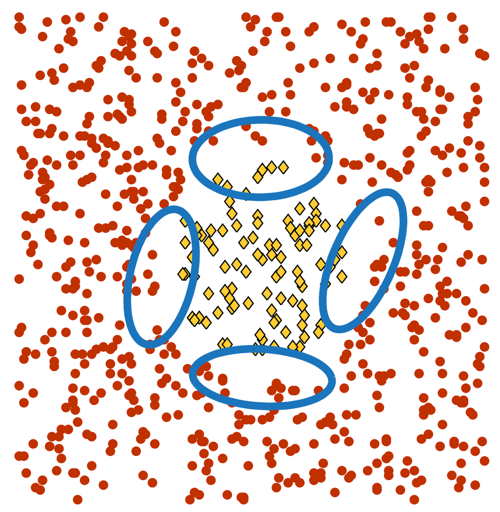
Theorem 3:
For squared exponential weights the derivative of the metric tensor is given as

$$\frac{\partial\mathrm{vec}\left[\mathbf{M}(c)\right]}{\partial c} = \frac{\rho}{\left(\sum_{j=1}^R \tilde{w}_j\right)^2}\sum_{r=1}^{R}\tilde{w}_r\mathrm{vec}\left[\mathbf{M}_r\right]\sum_{j=1}^{R}\tilde{w}_j\left((c-x_j)^T\mathbf{M}_j - (c-x_r)^T\mathbf{M}_r\right)$$

Geodesics as well as the exponential map can now be computed by solving the differential equations. We do this using standard numerical techniques.

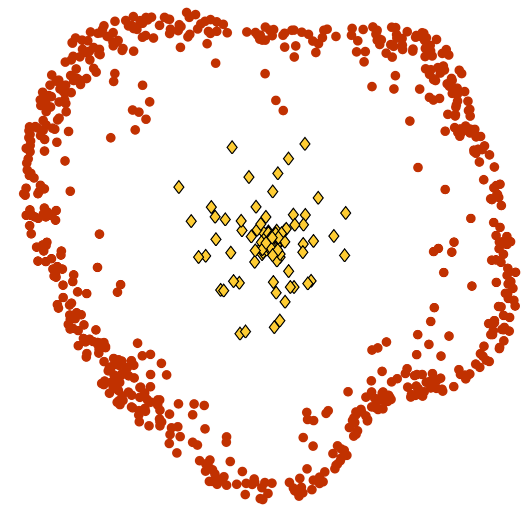
4 Results

Toy Example



Feature space & selected metric tensors.

- At random points in the feature space we compute a local LDA metric tensor [2].
- We combine tensors to form a manifold.
- We compute the mean and map the data to the tangent at the mean.
- We visualize the tangent space representation; note how classes are pushed apart.

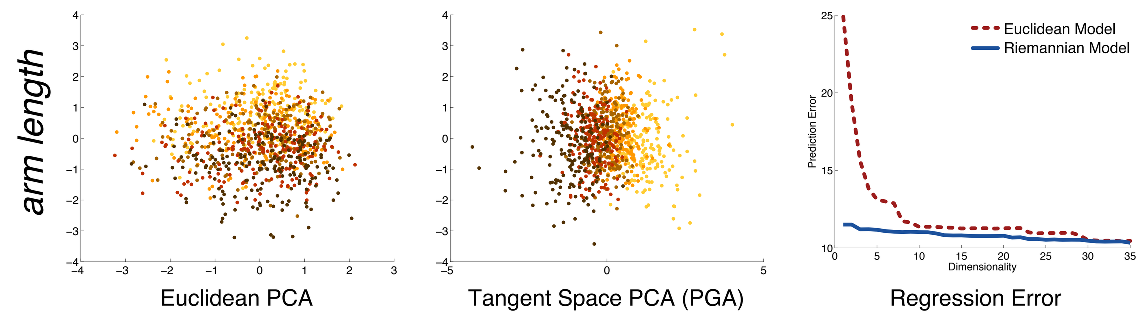


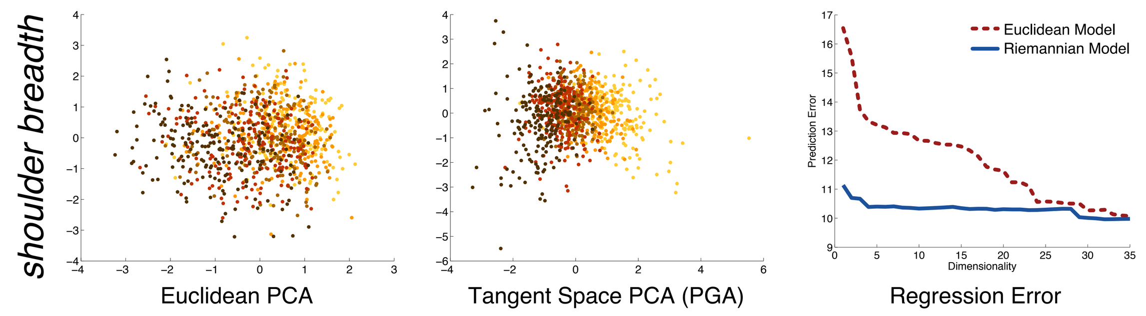
Tangent space representation.

Human Body Shapes

We seek representations of the human body that are related to measurements of the body (e.g. height, weight, etc.).

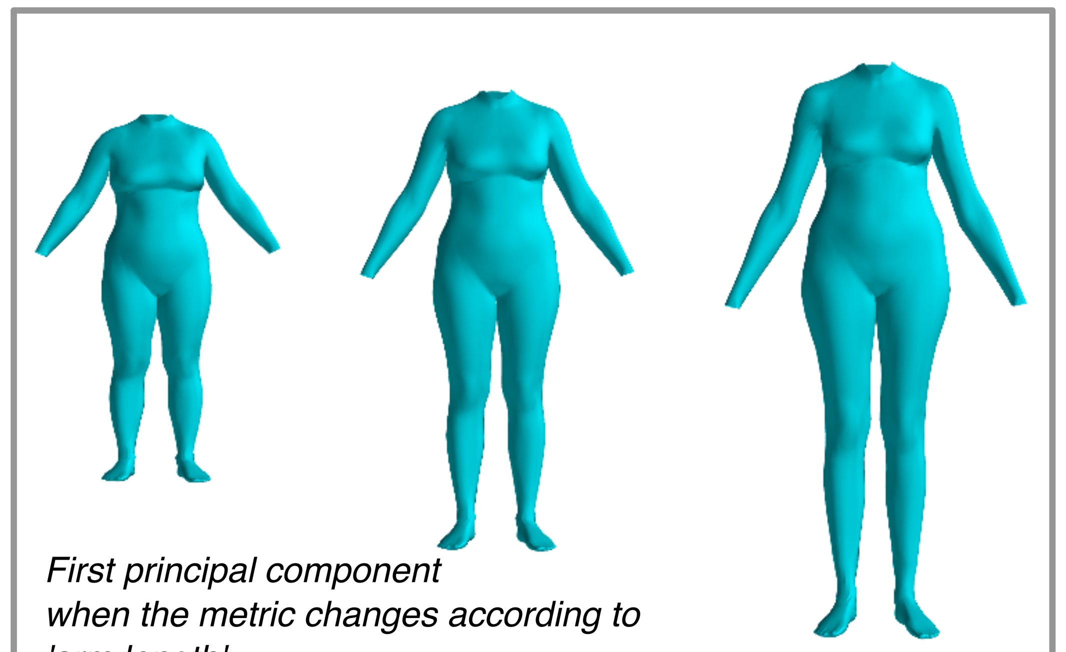
- Cluster the shapes according to a measurement;
- Learn a LMNN metric [3] for each cluster to push the clusters apart;
- Combine metrics to form a manifold and perform PCA on the manifold.



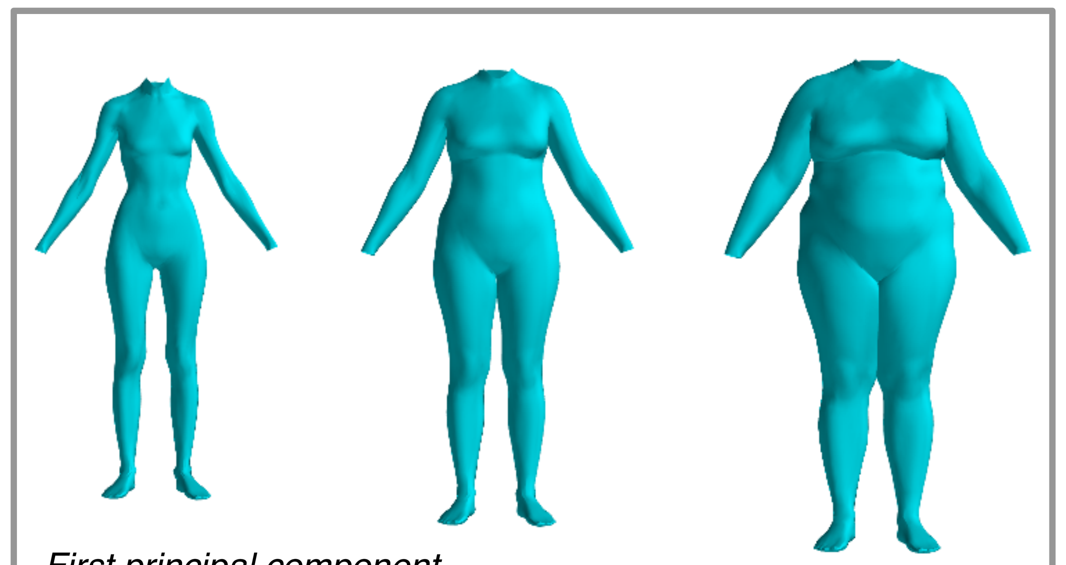


The plots show the tangent space representation and Euclidean PCA. Colors indicate clusters (according to the measurement).

We also perform regression in the tangent space. The plots show the regression error on a test set compared to the ordinary Euclidean approach.



First principal component when the metric changes according to 'arm length'.

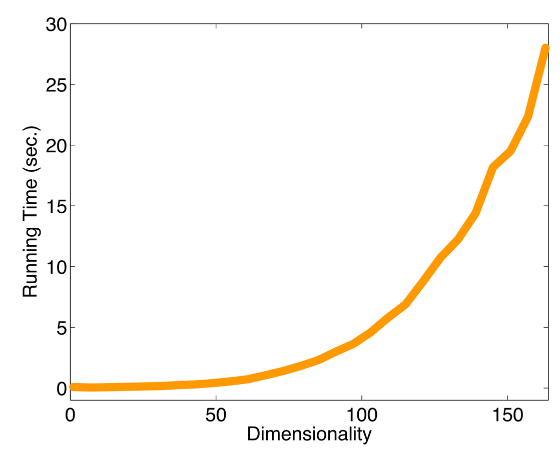


First principal component when the metric changes according to 'shoulder breadth'.

Running Time

Practical for feature spaces of less than 100 dimensions.

The plot shows running time for computing a geodesic.



Code is available online
`http://ps.is.tue.mpg.de/project/Smooth_Metric_Learning`

References

[1] S. Hauberg, O. Freifeld and M.J. Black, A Geometric take on Metric Learning, NIPS 2012

[2] T. Hastie and R. Tibshirani, Discriminative Adaptive Nearest Neighbor Classification, TPAMI, 1996

[3] K.Q. Weinberger and L.K. Saul, Distance Metric Learning for Large Margin Nearest Neighbor Classification, JMLR 2009

S.H. is supported in part by the Villum Foundation, and O.F. is supported in part by NIH-NINDS EUREKA (R01-NS066311)