Analysis of Three-Dimensional Shapes (IN2238, TU München, Summer 2014)

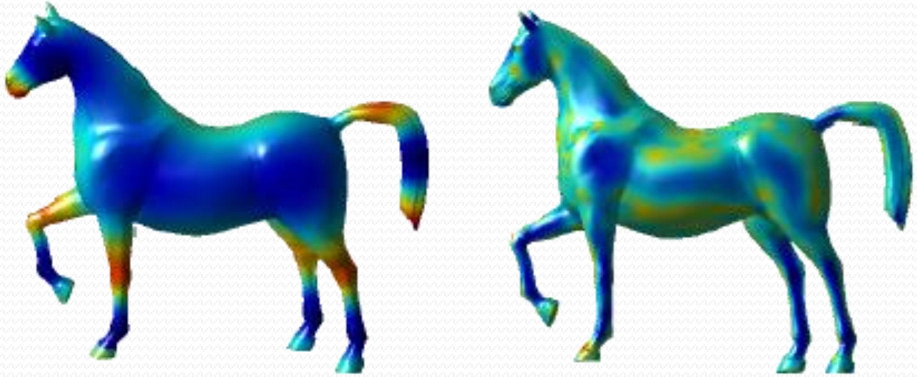
Euclidean Embeddings (05.05.2014)

Dr. Emanuele Rodolà
rodola@in.tum.de
Room 02.09.058, Informatik IX

Seminar

"Wave Kernel Signature"
Felix Müller

Wednesday, May 07
14:00 Room 02.09.023



Wrap-up

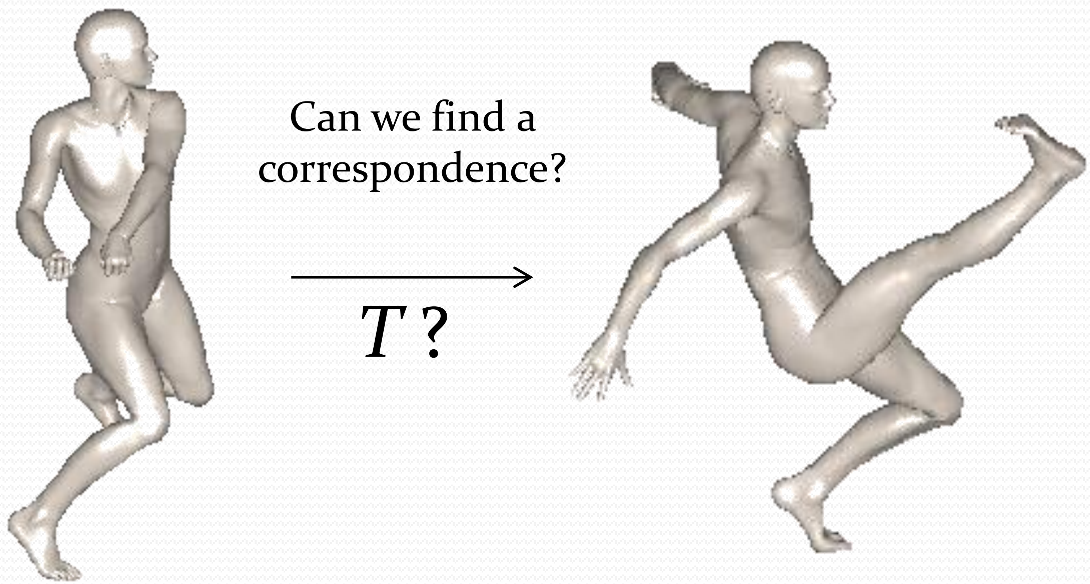
In the previous lectures, we have approached the problem of **shape similarity**…

$\overset{?}{\approx}$

Are the shapes similar?

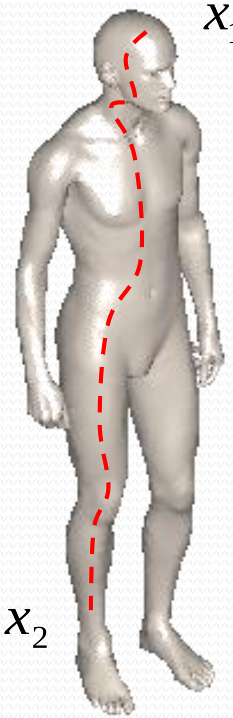
Wrap-up

In the previous lectures, we have approached the problem of **shape similarity**... and **shape matching**



Wrap-up

We modeled our shapes as **metric spaces**, that is, a set of points plus a metric (distance) function defined over it.



Wrap-up

We then asked if it is possible to define a meaningful notion of **distance** among these metric spaces.

It turned out that, *yes*, we can do that! In fact, there are several possible definitions:

- Lipschitz distance: measures the *relative* change of the metric
- Hausdorff distance: shapes are subsets of a common ambient space
- Gromov-Hausdorff distance: measures the *absolute* change of the metric

Wrap-up

We decided that the Gromov-Hausdorff distance captures the notion of shape similarity in the most natural way. Then we turned to the problem of actually **computing** this distance.

The **Gromov-Hausdorff distance** between two metric spaces X and Y is defined by

$$d_{\mathcal{GH}}(X,Y) = \inf_{Z,f,g} d_{\mathcal{H}}^{Z}(f(X), g(Y))$$

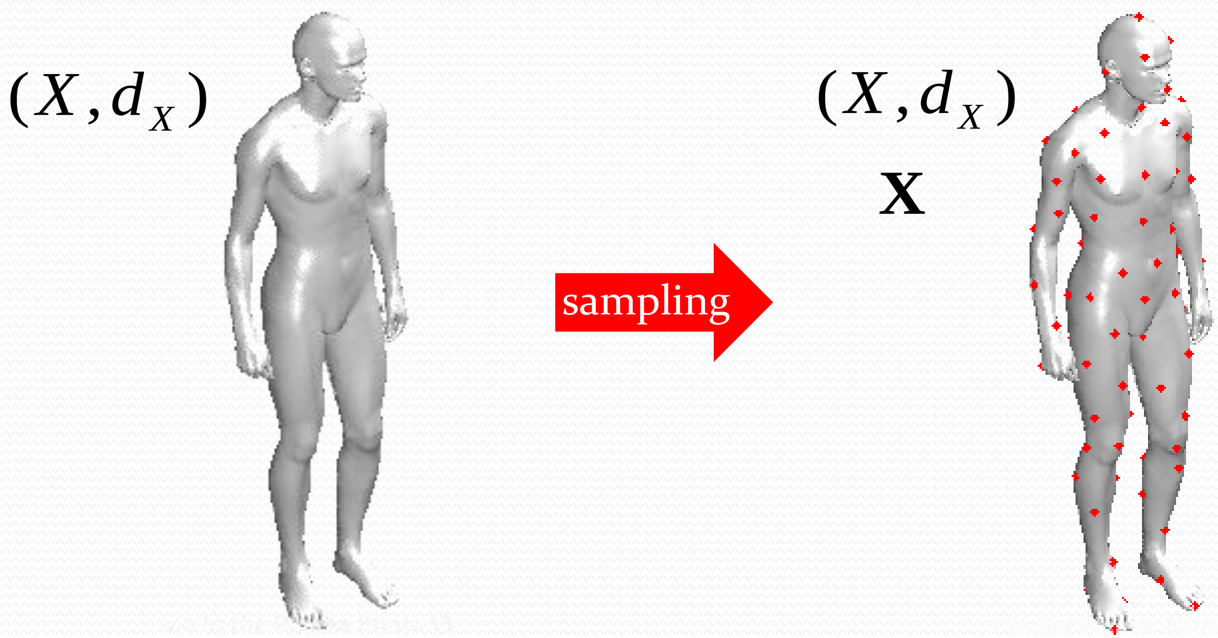
The infimum is taken over all ambient spaces Z and isometric embeddings (distance preserving) $f : X \to Z,\ \ g : Y \to Z$

$d_{\mathcal{GH}}$ is a metric on the space of isometry classes of compact metric spaces.

...how to do it?

Wrap-up

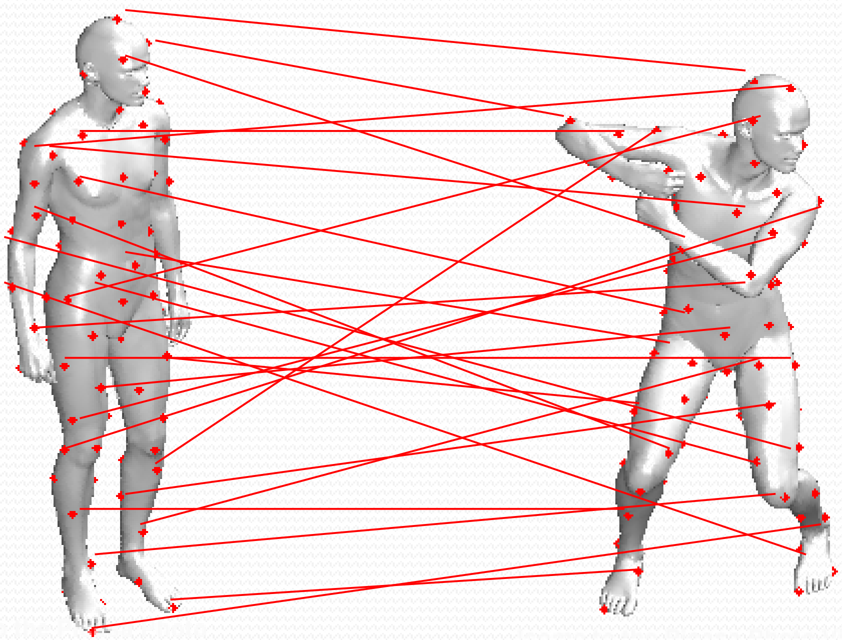
First, we have seen that restricting our attention to only the n **farthest point samples** of each shape still gives us a meaningful notion of distance.



Wrap-up

Then, we gave an alternative definition for the Gromov-Hausdorff distance, making use of the more intuitive notion of **correspondence**.

$$d_{\mathcal{GH}}(\mathbf{X},\mathbf{Y}) = \frac{1}{2} \inf_{R \subset \mathbf{X}\times\mathbf{Y}} \sup_{(x,y),(x',y')\in R} \left| d_X(x,x') - d_Y(y,y') \right|$$

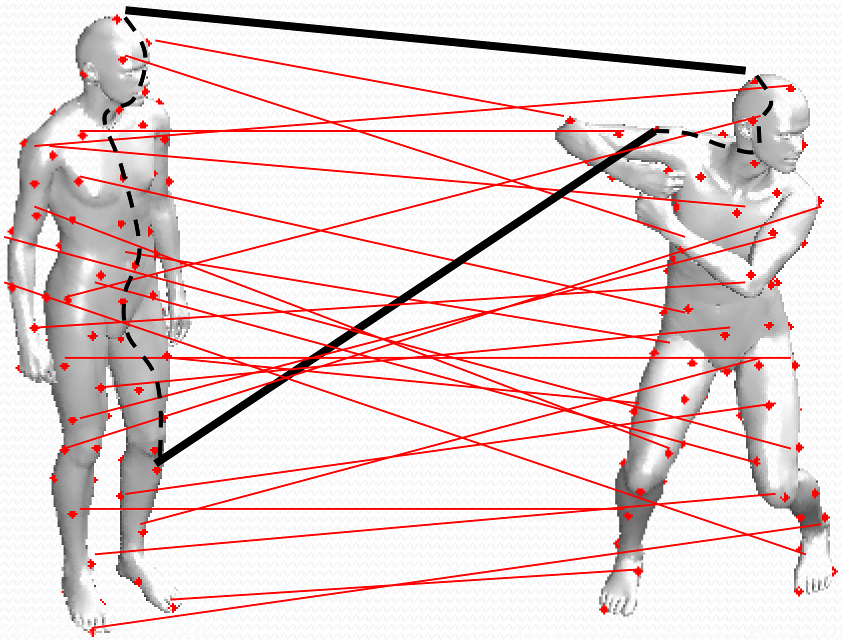


$$R = \{(x_1, y_3), (x_5, y_{12}), (x_2, y_7), (x_{10}, y_2), \ldots\}$$

Wrap-up

Then, we gave an alternative definition for the Gromov-Hausdorff distance, making use of the more intuitive notion of **correspondence**.

$$d_{\mathcal{GH}}(\mathbf{X},\mathbf{Y}) = \frac{1}{2} \inf_{R \subset \mathbf{X} \times \mathbf{Y}} \sup_{(x,y),(x',y') \in R} \left| d_X(x,x') - d_Y(y,y') \right|$$



$$R = \{(x_1, y_3), (x_5, y_{12}), (x_2, y_7), (x_{10}, y_2), \ldots\}$$

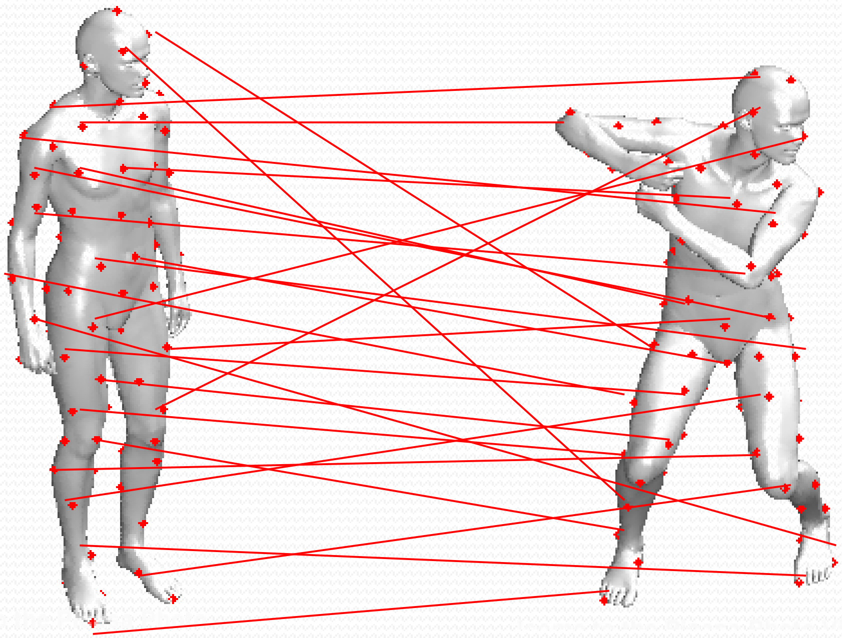
$$\Downarrow$$

$$\sup_{(x,y),(x',y') \in R} \left| d_X(x,x') - d_Y(y,y') \right| = 13.47$$

Wrap-up

Then, we gave an alternative definition for the Gromov-Hausdorff distance, making use of the more intuitive notion of **correspondence**.

$$d_{\mathcal{GH}}(\mathbf{X},\mathbf{Y}) = \frac{1}{2} \inf_{R \subset \mathbf{X}\times\mathbf{Y}} \sup_{(x,y),(x',y')\in R} \left| d_X(x,x') - d_Y(y,y') \right|$$

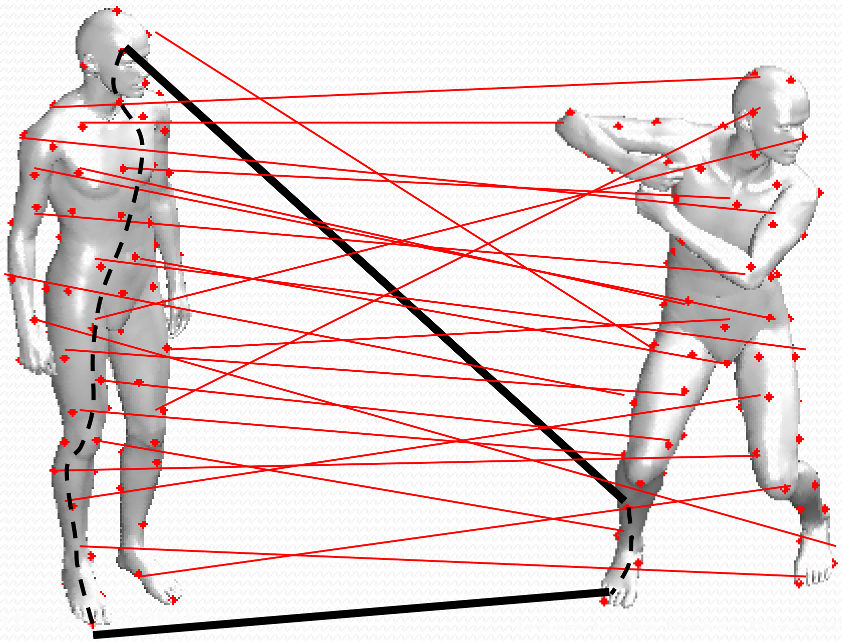


$$R = \{(x_1, y_{13}), (x_5, y_{23}), (x_2, y_2), (x_{10}, y_9), \ldots\}$$

Wrap-up

Then, we gave an alternative definition for the Gromov-Hausdorff distance, making use of the more intuitive notion of **correspondence**.

$$d_{\mathcal{GH}}(\mathbf{X},\mathbf{Y}) = \frac{1}{2} \inf_{R \subset \mathbf{X}\times\mathbf{Y}} \sup_{(x,y),(x',y')\in R} \left| d_X(x,x') - d_Y(y,y') \right|$$



$$R = \{(x_1, y_{13}), (x_5, y_{23}), (x_2, y_2), (x_{10}, y_9), \ldots\}$$

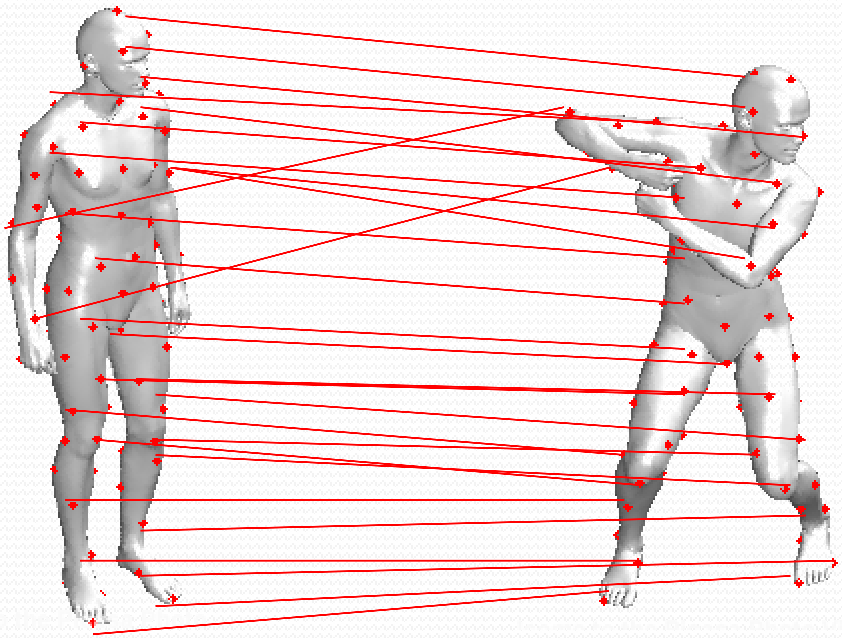
$$\Downarrow$$

$$\sup_{(x,y),(x',y')\in R} \left| d_X(x,x') - d_Y(y,y') \right| = 21.14$$

Wrap-up

Then, we gave an alternative definition for the Gromov-Hausdorff distance, making use of the more intuitive notion of **correspondence**.

$$d_{\mathcal{GH}}(\mathbf{X},\mathbf{Y}) = \frac{1}{2} \inf_{R\subset \mathbf{X}\times\mathbf{Y}} \sup_{(x,y),(x',y')\in R} \left| d_X(x,x') - d_Y(y,y') \right|$$

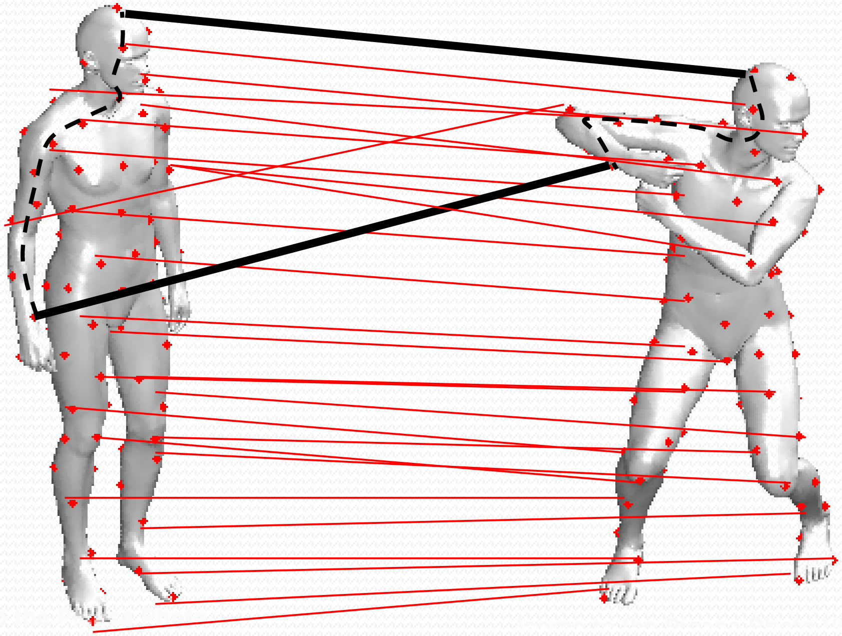


$$R = \{(x_1, y_{23}), (x_5, y_{41}), (x_2, y_7), (x_{10}, y_{19}), \ldots\}$$

Wrap-up

Then, we gave an alternative definition for the Gromov-Hausdorff distance, making use of the more intuitive notion of **correspondence**.

$$d_{\mathcal{GH}}(\mathbf{X},\mathbf{Y}) = \frac{1}{2} \inf_{R \subset \mathbf{X} \times \mathbf{Y}} \sup_{(x,y),(x',y') \in R} \left| d_X(x,x') - d_Y(y,y') \right|$$



$$R = \{(x_1, y_{23}), (x_5, y_{41}), (x_2, y_7), (x_{10}, y_{19}), \ldots\}$$

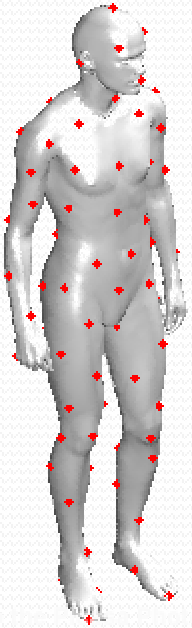
$$\Downarrow$$

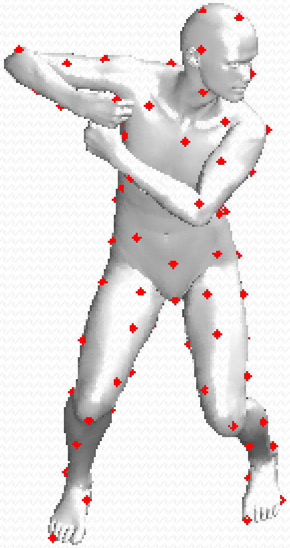
$$\sup_{(x,y),(x',y') \in R} \left| d_X(x,x') - d_Y(y,y') \right| = 3.07$$

Wrap-up

Then, we gave an alternative definition for the Gromov-Hausdorff distance making use of the more intuitive notion of **correspondence**.

$$d_{\mathcal{GH}}(\mathbf{X},\mathbf{Y}) = \frac{1}{2} \inf_{R \subset \mathbf{X} \times \mathbf{Y}} \sup_{(x,y),(x',y') \in R} \left| d_X(x,x') - d_Y(y,y') \right|$$





Requires considering all possible correspondences. There are $n!$ of them ☹

Wrap-up

Passing to **matrix notation**, we wrote:

$$\frac{1}{2}\inf_{R\subset \mathbf{X}\times\mathbf{Y}}\sup_{(x,y),(x',y')\in R}\left|d_X(x,x')-d_Y(y,y')\right|$$

$$\Downarrow$$

$$\frac{1}{2}\min_{R}\max_{i,j,\ell,m} C_{(i\ell)(jm)}R_{ij}R_{\ell m}$$

where R is a binary *correspondence matrix* and C is a *cost matrix*.

Wrap-up

Since the original problem seems difficult to solve, we had a look at a few possible **relaxations**. A "relaxation" is an approximation of a difficult problem by another similar problem that is easier to solve.

Hopefully, the solution to the relaxed problem will provide some information about the original solution.

$$\underset{\text{difficult}}{\min_x f(x)} \xrightarrow{\text{relax}} \underset{\text{easy}}{\min_y g(y)} \xrightarrow{\text{solve}}$$

Interpret the solution to the relaxed problem as a solution to the original one

Wrap-up

First relaxation: relax the max to a sum.

$$\frac{1}{2}\min_{R}\max_{i,j,\ell,m} C_{(i\ell)(jm)}R_{ij}R_{\ell m}$$

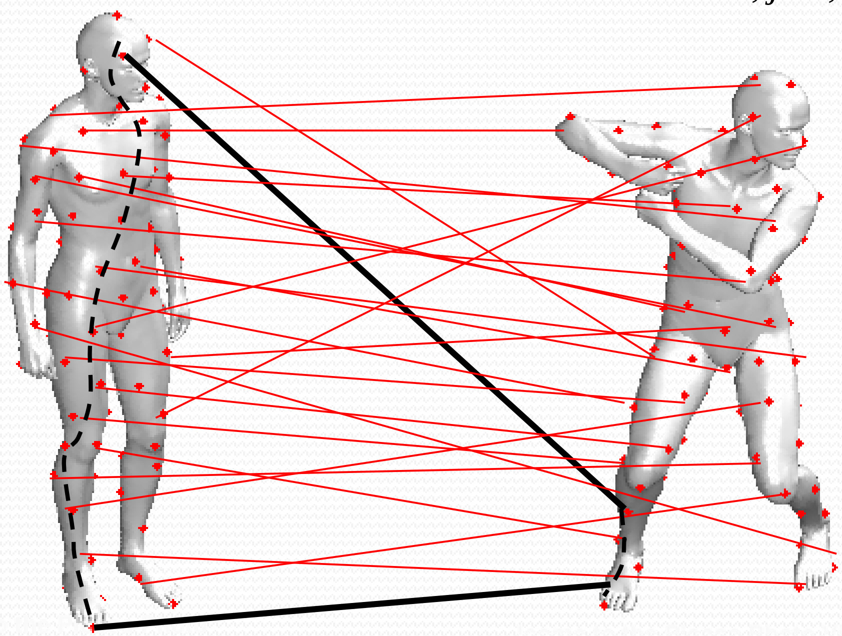
$$\Downarrow$$

$$\frac{1}{2}\min_{R}\sum_{i,j}\sum_{\ell,m} C_{(i\ell)(jm)}R_{ij}R_{\ell m}$$

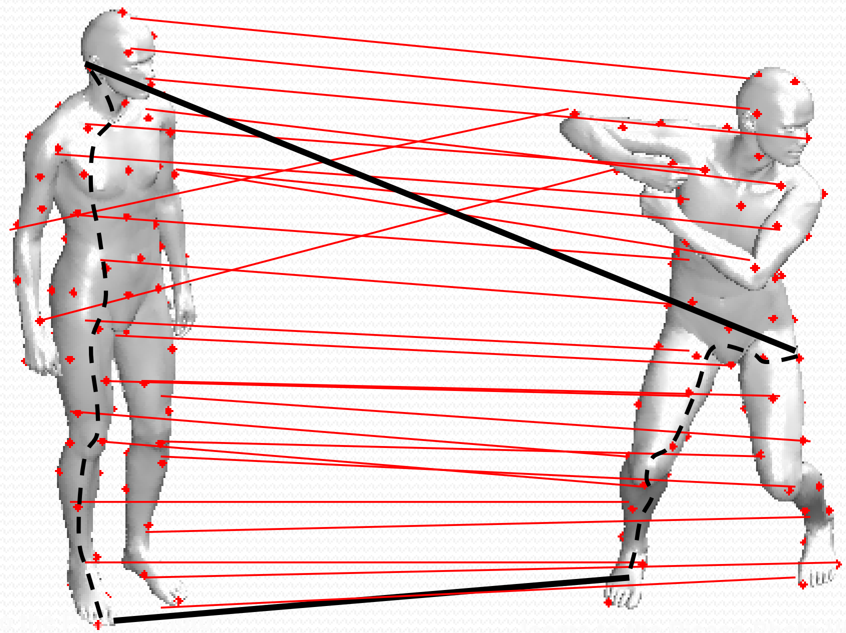
Wrap-up

First relaxation: replace the max with a sum.

$$\frac{1}{2}\min_{R}\sum_{i,j}\sum_{\ell,m} C_{(i\ell)(jm)} R_{ij} R_{\ell m}$$



$\text{max} = 21.14 \quad \text{sum} = 207.44$



$\text{max} = 27.48 \quad \text{sum} = 41.02$

Wrap-up

Simplify using the dreadful matrix notation.

$$\frac{1}{2}\min_{R}\sum_{i,j}\sum_{\ell,m} C^{(p)}_{(i\ell)(jm)} R_{ij} R_{\ell m}$$

?

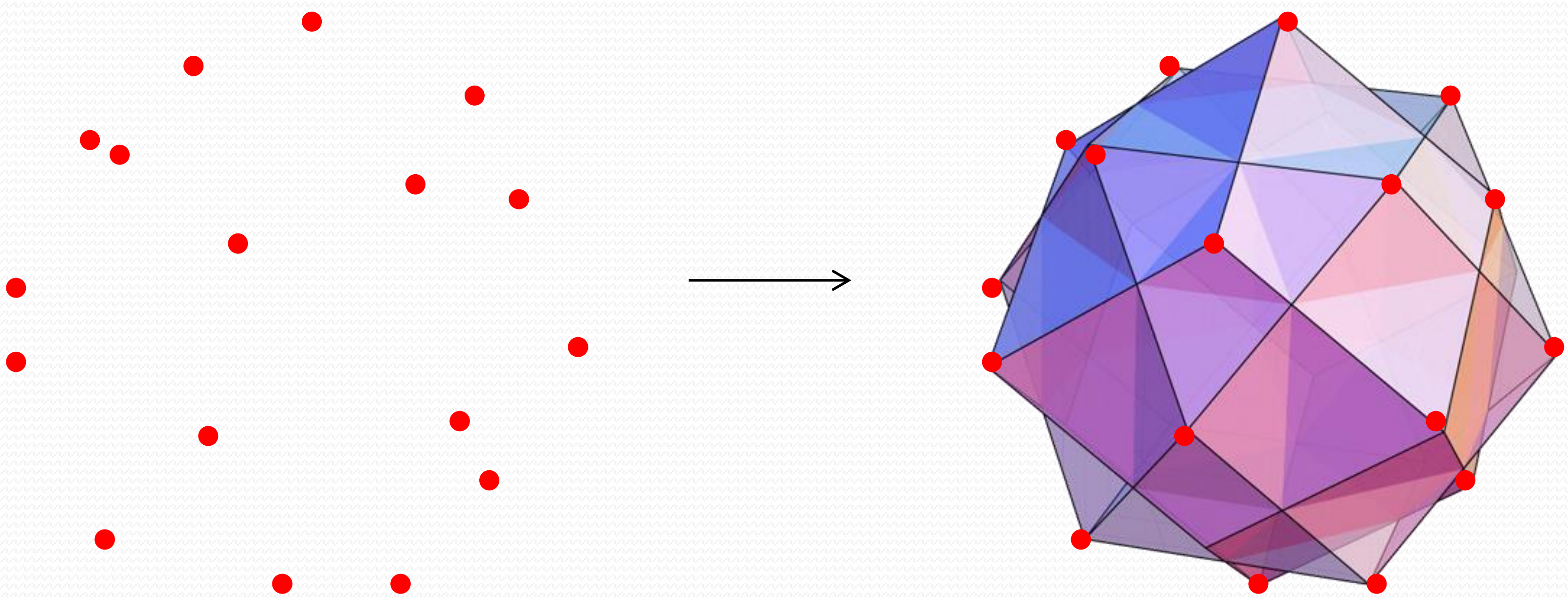
$$\min_{R\in\{0,1\}^{n\times n}} \operatorname{vec}\{R\}^{\mathrm{T}} C \operatorname{vec}\{R\}$$

$$\text{s.t.} \quad R\mathbf{1}=\mathbf{1},\ \ R^{\mathrm{T}}\mathbf{1}=\mathbf{1}$$

Wrap-up

Second relaxation: replace binary solutions with continuous solutions.

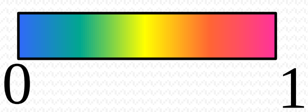
$$R \in \{0,1\}^{n \times n} \rightarrow R \in [0,1]^{n \times n}$$

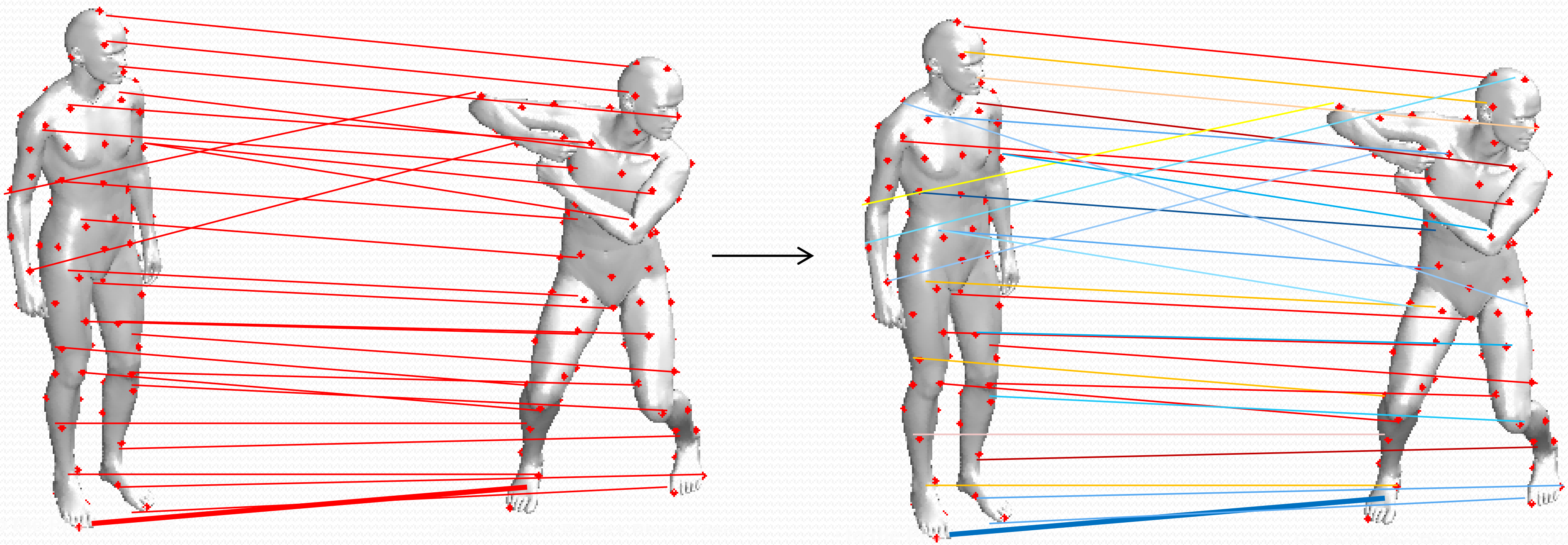


Wrap-up

Second relaxation: replace binary solutions with continuous solutions.

$$R \in \{0,1\}^{n\times n} \rightarrow R \in [0,1]^{n\times n}$$





Wrap-up

Other relaxations: replace the mapping constraints…

$$R\mathbf{1} = \mathbf{1}, \;\; R^{\mathrm{T}}\mathbf{1} = \mathbf{1}$$

$$\|R\|^2 = 1$$

$$\mathbf{1}^{\mathrm{T}} R\mathbf{1} = 1$$

…and / or replace the cost function

$$C_{(i\ell)(jm)} = \left| d_{\mathbf{X}}(x_i, x_j) - d_{\mathbf{Y}}(y_\ell, y_m) \right|^p$$

$$C_{(i\ell)(jm)} = e^{-\beta \left| d_{\mathbf{X}}(x_i, x_j) - d_{\mathbf{Y}}(y_\ell, y_m) \right|^2}$$

Wrap-up

Conclusions: The Gromov-Hausdorff distance is difficult to compute, but it can be approximated by efficient relaxations.

Some of these are also effective, but the whole approach in general suffers from the following drawbacks:

- The approach does not scale well with the size of the shapes
- The connection with Gromov-Hausdorff gets lost easily
- It is difficult to give guarantees on the quality of the solution

Hausdorff revisited

Recall that the Gromov-Hausdorff distance is defined in terms of the **Hausdorff** distance:

$$d_{\mathcal{GH}}(X,Y) = \inf_{Z,f,g} d_{\mathcal{H}}^{Z}(f(X), g(Y))$$

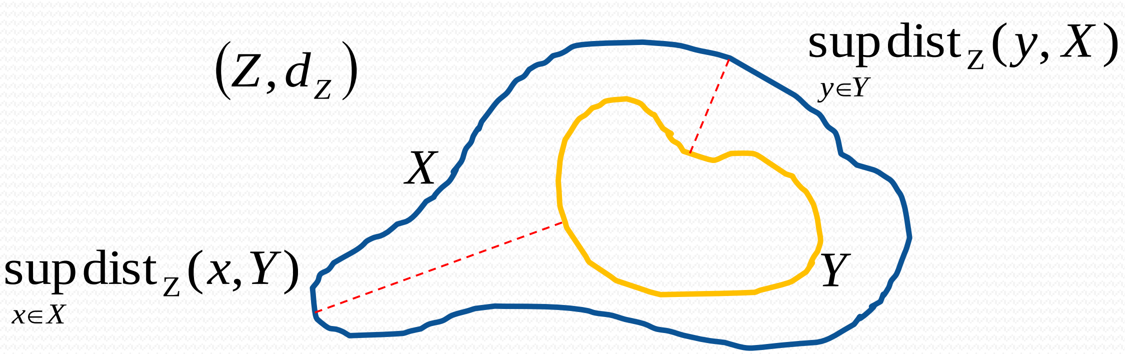
where $f : X \to Z,\ \ g : Y \to Z$ are isometric embeddings

Hausdorff revisited

Recall that the Gromov-Hausdorff distance is defined in terms of the **Hausdorff** distance:

The **Hausdorff distance** between two **compact subsets** $X, Y \subset (Z, d_Z)$ is defined by

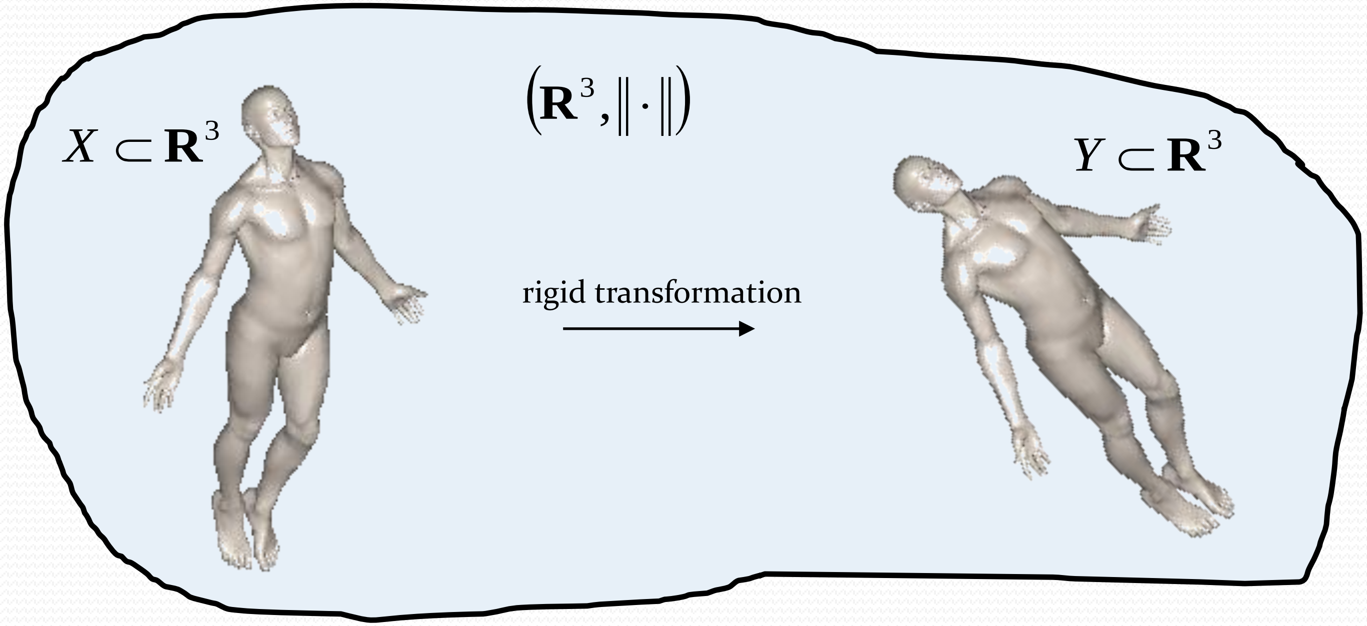
$$d_{\mathcal{H}}^{Z}(X,Y) = \max\left\{ \sup_{x\in X} \mathrm{dist}_Z(x,Y), \sup_{y\in Y} \mathrm{dist}_Z(y,X) \right\}$$



Hausdorff revisited

The **Hausdorff distance** between two **compact subsets** $X, Y \subset (Z, d_Z)$ is defined by

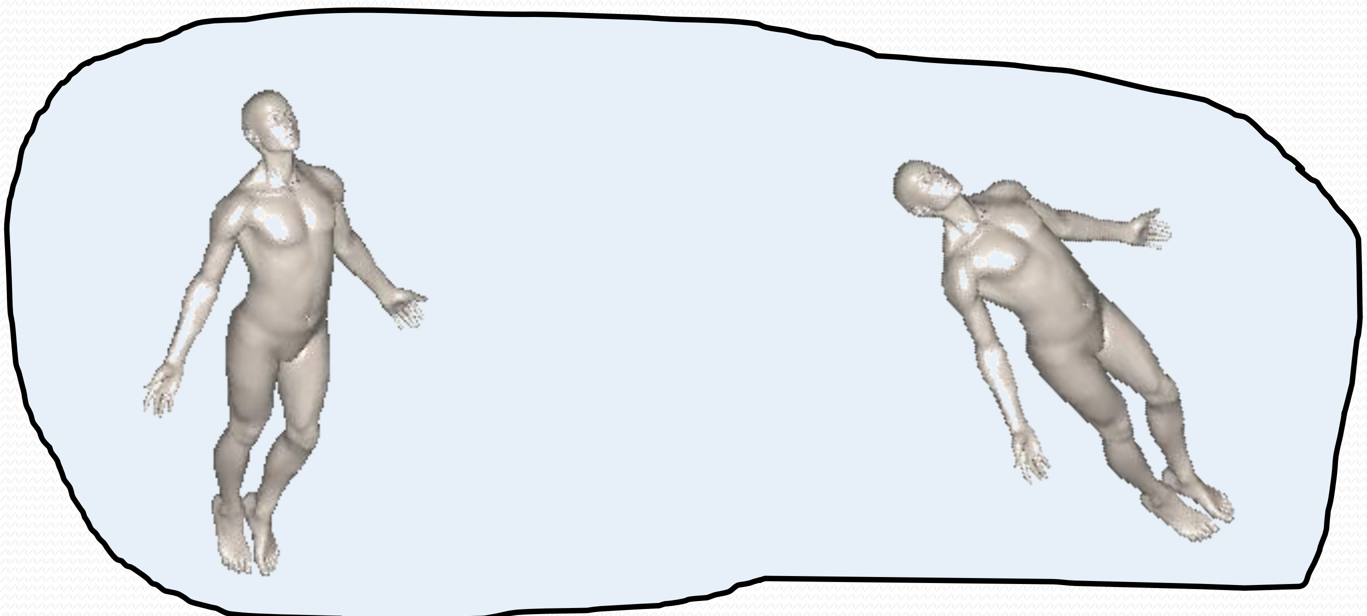
$$d_{\mathcal{H}}^{Z}(X,Y) = \max\left\{\sup_{x \in X} \mathrm{dist}_Z(x,Y), \sup_{y \in Y} \mathrm{dist}_Z(y,X)\right\}$$



Hausdorff revisited

The **Hausdorff distance** between two **compact subsets** $X, Y \subset (Z, d_Z)$ is defined by

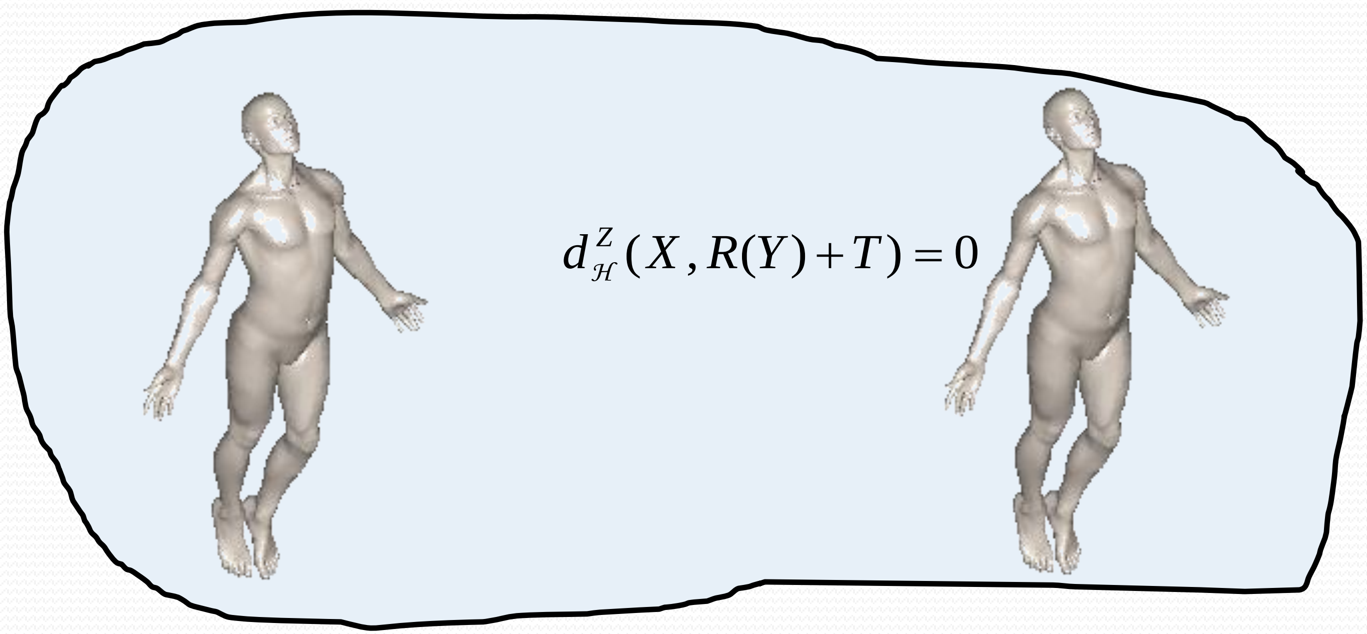
$$d_{\mathcal{H}}^{Z}(X,Y) = \max\left\{ \sup_{x \in X} \mathrm{dist}_Z(x,Y), \sup_{y \in Y} \mathrm{dist}_Z(y,X) \right\}$$



Hausdorff revisited

The **Hausdorff distance** between two **compact subsets** $X, Y \subset (Z, d_Z)$ is defined by

$$d_{\mathcal{H}}^{Z}(X,Y) = \max\left\{ \sup_{x \in X} \mathrm{dist}_Z(x,Y), \sup_{y \in Y} \mathrm{dist}_Z(y,X) \right\}$$



Can be solved e.g. via ICP

Euclidean embeddings

Let's have a look at the Gromov-Hausdorff distance again:

$$d_{\mathcal{GH}}(X,Y) = \inf_{Z,f,g} d_{\mathcal{H}}^{Z}(f(X), g(Y))$$

where $f: X \to Z$, $g: Y \to Z$ are isometric embeddings

From the previous example we have seen that optimizing for rigid transformations is much simpler and many effective algorithms exist (ICP).

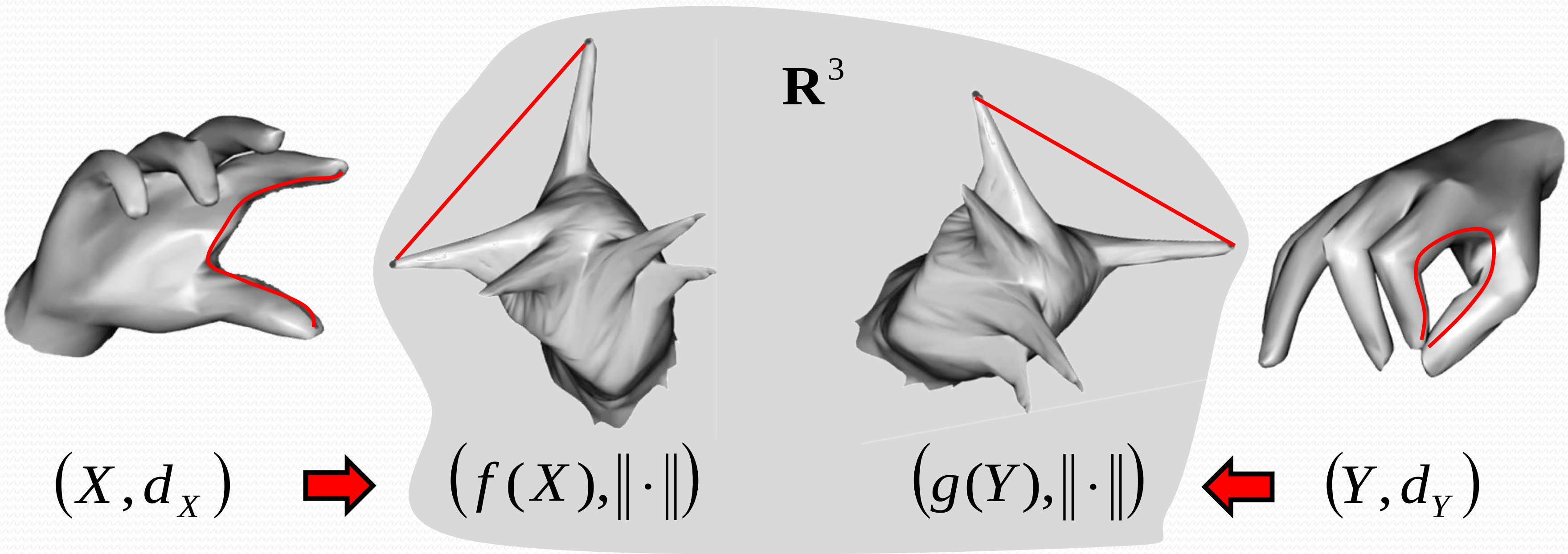
Then can't we just map each shape to $\mathbf{R}^3$ and then solve the resulting **rigid** problem there?

Euclidean embeddings

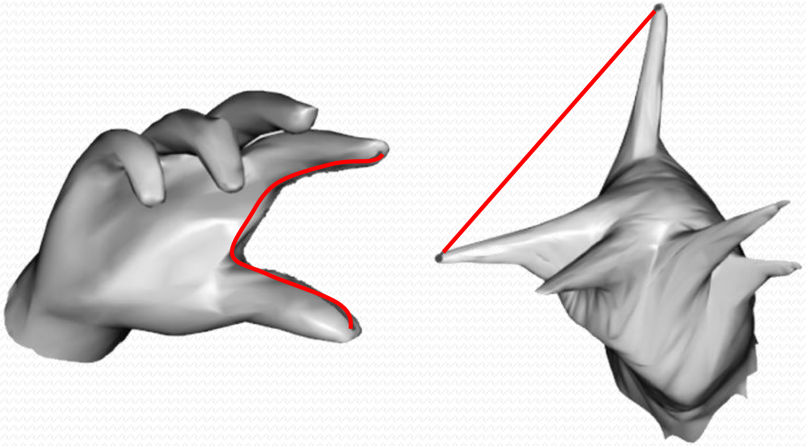
$$d_{\mathcal{GH}}(X,Y) = \inf_{\mathbf{R}^3, f, g} d_{\mathcal{H}}^{\mathbf{R}^3}(f(X), g(Y))$$

where $f : X \to \mathbf{R}^3$, $g : Y \to \mathbf{R}^3$ are isometric embeddings

In other words, we are looking for something like:



Euclidean embeddings



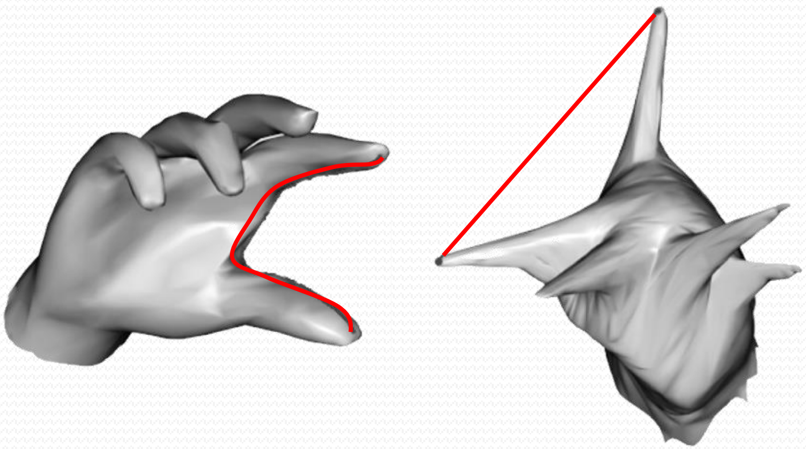
$$\left(X, d_X\right) \Rightarrow \left(f(X), \|\cdot\|\right)$$

Thus, we would like to find a map $f:(X, d_X) \to (\mathbf{R}^m, \|\cdot\|)$ such that

$$d_X(x, x') = \|f(x) - f(x')\|_2$$

for all $x, x' \in X$

Euclidean embeddings



$$(X, d_X) \quad \Rightarrow \quad \left(f(X), \|\cdot\|\right)$$

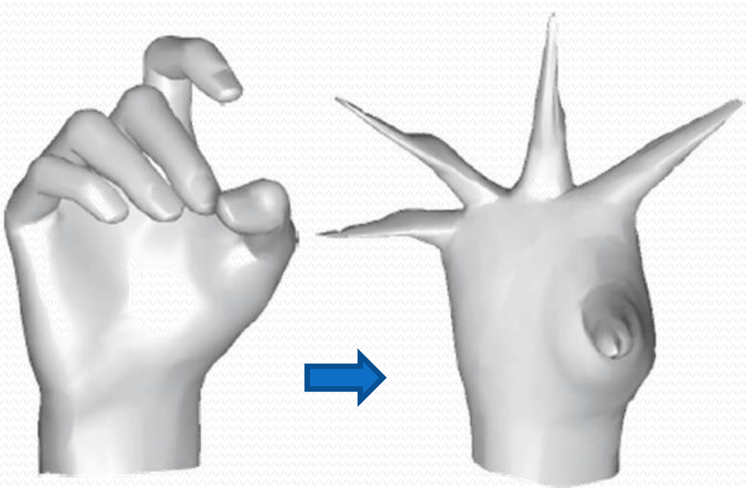
The image $f(X)$ is also called the *canonical form* of X.

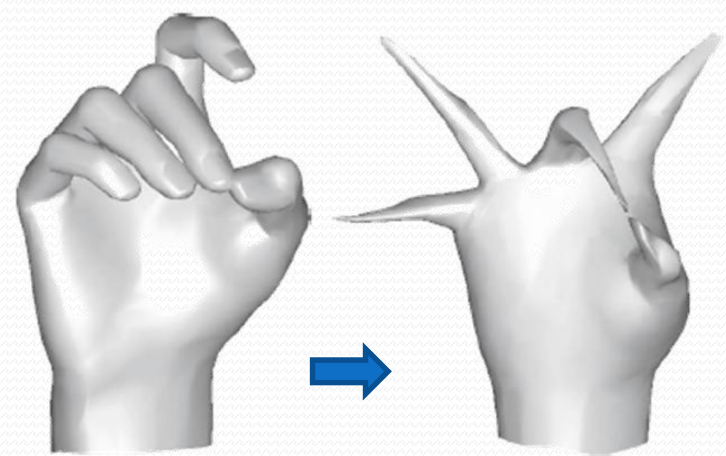
It defines an equivalence class of shapes up to an isometry in $\mathbf{R}^m$ (these correspond to rotations, translations, reflections).

Euclidean embeddings

Note that:

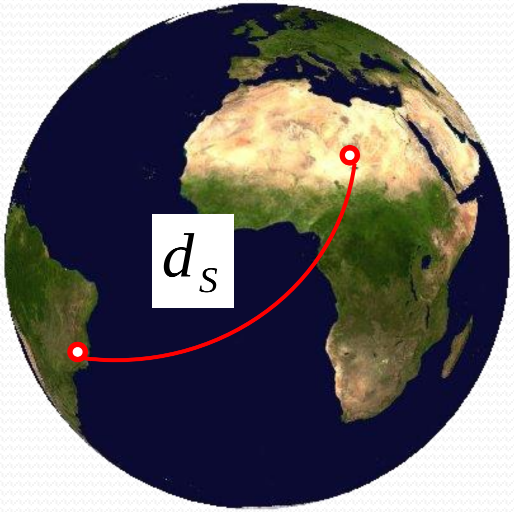
- We are assuming *m* to be arbitrary (i.e. not necessarily *m*=3). This will allow us to keep the approach general, and we will need it in the near future.
- Topological noise can significantly alter distances.

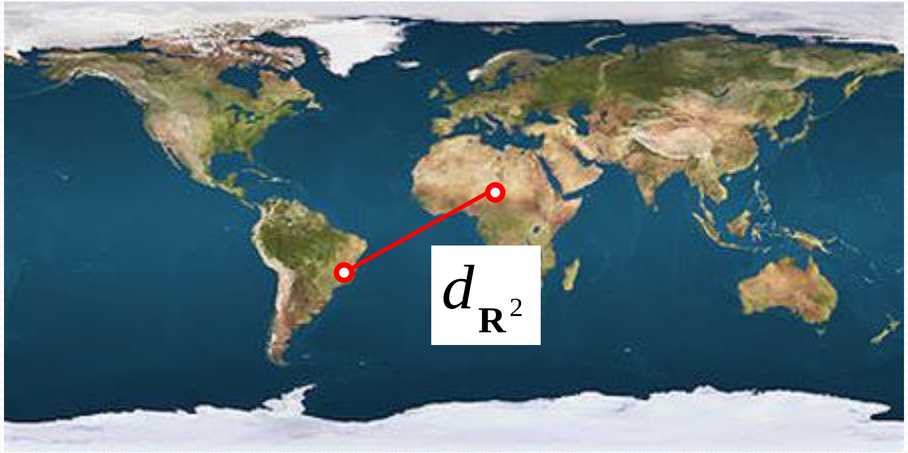




A cartographer's problem

- We still don't know to what extent our shapes X are «isometrically embeddable» into $\mathbf{R}^m$!

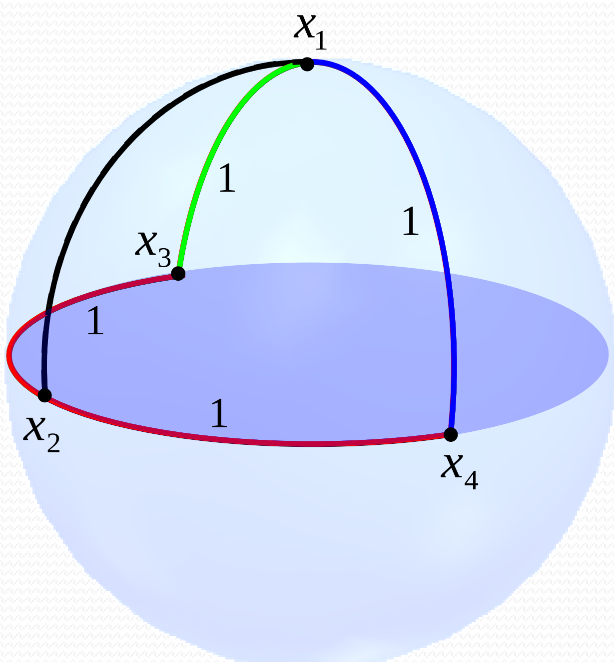




$$d_S \stackrel{?}{=} d_{\mathbf{R}^2}$$

Impossible to do without introducing distortion.

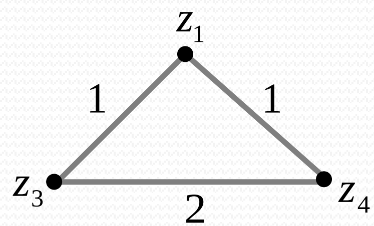
The smallest non-trivial example



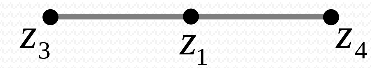
$$D_S = \begin{array}{c} \\ \\ \end{array}\begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \end{array}$$

$$D_S = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 2 \\ 1 & 1 & 2 & 0 \end{pmatrix} \begin{array}{c} x_1 \\ x_2 \\ x_3 \\ x_4 \end{array}$$

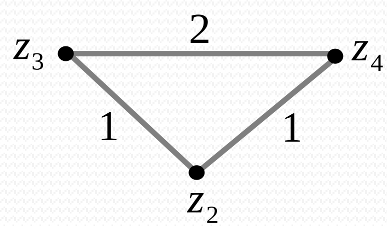
Assume $(D_S)_{ij} = d_{\mathbf{R}^m}(z_i, z_j)$ and consider the triangle z_3, z_1, z_4



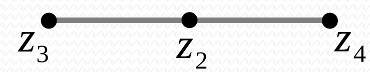
collinear!



Now consider the triangle z_3, z_2, z_4



collinear!



Then $z_1 = z_2$, which contradicts $(D_S)_{12} = d_{\mathbf{R}^m}(z_1, z_2) = 1$

This metric space cannot be embedded into a Euclidean space of *any* finite dimension!

Minimum-distortion embedding

Still, we could try to look for an approximate embedding, such that the distortion of d_X is mimimal according to some criterion.

One such criterion is the usual metric distortion induced by the mapping *f*:

$$\operatorname{dis} f = \sup_{x_i, x_j \in X} \left| d_X(x_i, x_j) - d_{\mathbf{R}^m}(f(x_i), f(x_j)) \right|$$

A *minimum-distortion embedding* would then be the *f* minimizing the above.

Minimum-distortion embedding

$$\operatorname{dis} f = \sup_{x_i, x_j \in X} \left| d_X(x_i, x_j) - d_{\mathbf{R}^m}(f(x_i), f(x_j)) \right|$$

We can define alternative measures of distortion as well, for instance:

$$\sigma_p(f) = \sum_{i>j} \left| d_X(x_i, x_j) - d_{\mathbf{R}^m}(f(x_i), f(x_j)) \right|^p$$

Keep in mind the the resulting canonical form $f(X)$ will only be an approximation. <u>The embedding introduces a distortion</u>, which in turn influences the accuracy of our similarity calculations.

Minimum-distortion embedding

We will consider the *quadratic stress* $\sigma_2(f)$. Then we would like to compute:

$$f = \arg\min_{f: X \to \mathbf{R}^m} \sum_{i>j} \left| d_X(x_i, x_j) - d_{\mathbf{R}^m}(f(x_i), f(x_j)) \right|^2$$

Let us consider a sampling $\{x_1, \ldots, x_N\}$ of N points over X, and denote their images as $z_i = f(x_i)$. Arranging the z_i into a $N \times m$ matrix $Z = (z_i^j)$, we can rewrite the distortion criterion as

$$\sigma_2(Z, D_X) = \sum_{i>j} \left| d_X(x_i, x_j) - d_{ij}(Z) \right|^2$$

where $d_{ij}(Z) = \left\| z_i - z_j \right\|_2$

Differently from the matching problem, now Z is the unknown!

Minimum-distortion embedding

$$Z^* = \underset{Z \in \mathbf{R}^{N \times m}}{\arg\min}\, \sigma_2(Z)$$

Note that there is no unique solution, in fact applying any Euclidean isometry to Z^* will not change the value of σ_2 .

Problems of this sort started appearing in psychology in the 1950's, and are usually referred to as **multidimensional scaling** (MDS) problems.

PSYCHOMETRIKA—VOL. 27, NO. 2
JUNE, 1962

THE ANALYSIS OF PROXIMITIES: MULTIDIMENSIONAL SCALING WITH AN UNKNOWN DISTANCE FUNCTION. I.

ROGER N. SHEPARD

BELL TELEPHONE LABORATORIES

Multidimensional scaling

Empirical procedures of several diverse kinds have this in common: they start with a fixed set of entities and determine, for every pair of these, a number reflecting how closely the two entities are related psychologically. The nature of the psychological relation depends upon the nature of the entities. If the entities are all stimuli or all responses, we are inclined to think of the relation as one of similarity. A somewhat more objective (though less intuitive) characterization of such a relation, perhaps, is that of substitutability. The statement that stimulus A is more similar to B than to C, for example, could be interpreted to say that the psychological (or behavioral) consequences are greater when C, rather than B, is substituted for A. From this standpoint a natural procedure for determining similarities of stimuli or responses is by recording substitution errors during identification learning [2, 7, 12, 14, 17, 18]. In addition, though, disjunctive reaction time and sorting time have also been proposed as measures of psychological similarity [20]. Finally, of course, individuals have sometimes been instructed simply to rate each pair of stimuli, directly, on a scale of apparent similarity [1, 6]. The notion of similarity is not necessarily restricted to stimuli or responses (in the narrow sense of these words), however. Serviceable measures of similarity may also be found for concepts, attitudes, personality structures, or even social institutions, political systems, and the like.

Quadratic stress

$$\sigma_2(Z, D_X) = \sum_{i>j} \left| d_X(x_i, x_j) - d_{ij}(Z) \right|^2$$

For any given configuration Z, the **stress** measures how well that configuration matches the data. We look for the configuration of minimum stress.

Let's rewrite the stress function differently:

$$\begin{aligned}\sigma_2(Z, D_X) &= \sum_{i>j} \left| d_X(x_i, x_j) - d_{ij}(Z) \right|^2 \\ &= \sum_{i>j} \underbrace{d_{ij}^2(Z)}_{\text{Term 1}} - \underbrace{2d_{ij}(Z)\, d_X(x_i, x_j)}_{\text{Term 2}} + d_X^2(x_i, x_j)\end{aligned}$$

Quadratic stress (Term 1)

z_i is the i-th row of Z and z_i^k is its k-th coordinate

$$\sum_{i>j} d_{ij}^2(Z) =$$

$$= \sum_{i>j} \left\| z_i - z_j \right\|^2 = \sum_{i>j} \sum_{k=1}^{m} (z_i^k - z_j^k)^2$$

$$= \sum_{i>j} \sum_{k=1}^{m} \left(z_i^k\right)^2 - 2 z_i^k z_j^k + \left(z_j^k\right)^2$$

$$= \sum_{i>j} \langle z_j, z_j \rangle + \langle z_i, z_i \rangle - 2\langle z_i, z_j \rangle = \sum_{i>j} \langle z_j, z_j \rangle + \langle z_i, z_i \rangle \boxed{- 2 \sum_{i>j} \langle z_i, z_j \rangle}$$

$$= (N-1) \sum_{i=1}^{N} \langle z_i, z_i \rangle \boxed{- \left(\sum_{i,j} \langle z_i, z_j \rangle - \sum_{i=1}^{N} \langle z_i, z_i \rangle \right)}$$

$$= N \sum_{i=1}^{N} \langle z_i, z_i \rangle - \sum_{i,j} \langle z_i, z_j \rangle$$

Quadratic stress (Term 1)

$$\sum_{i>j} d_{ij}^2(Z) = N\sum_{i=1}^{N}\langle z_i, z_i\rangle - \sum_{i,j}\langle z_i, z_j\rangle$$

$$= N\,\mathrm{tr}(ZZ^{\mathrm{T}}) - \mathrm{tr}(1_{N\times N}ZZ^{\mathrm{T}})$$

$1_{N\times N}$ is a matrix of ones

$$= \mathrm{tr}(VZZ^{\mathrm{T}})$$

$$V_{ij} = \begin{cases} -1 & i \neq j \\ N-1 & i = j \end{cases}$$

$$= \mathrm{tr}(Z^{\mathrm{T}}VZ)$$

The last step can be done because

$$\mathrm{tr}(AB) = \sum_{i=1}^{m}(AB)_{ii} = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}B_{ji} = \sum_{j=1}^{n}\sum_{i=1}^{m} B_{ji}A_{ij} = \sum_{j=1}^{n}(BA)_{jj} = \mathrm{tr}(BA)$$

Quadratic stress (Term 2)

$$\sigma_2(Z,D_X)=\underbrace{\sum_{i>j} d_{ij}^2(Z)}_{\mathrm{tr}(Z^{\mathrm{T}}VZ)}\underbrace{-2d_{ij}(Z)d_X(x_i,x_j)}_{\text{(let's do it)}}+d_X^2(x_i,x_j)$$

$$\begin{aligned}\sum_{i>j} d_{ij}(Z)d_X(x_i,x_j) &= \sum_{i>j} d_X(x_i,x_j)d_{ij}^{-1}(Z)d_{ij}^2(Z)\\ &= \sum_{i>j} d_X(x_i,x_j)d_{ij}^{-1}(Z)\sum_{k=1}^{m}(z_i^k-z_j^k)^2\\ &= \sum_{i>j} d_X(x_i,x_j)d_{ij}^{-1}(Z)\Big(\langle z_i,z_i\rangle+\langle z_j,z_j\rangle-2\langle z_i,z_j\rangle\Big)\end{aligned}$$

Quadratic stress (Term 2)

$$\sum_{i>j} d_{ij}(Z)\, d_X(x_i,x_j) = \sum_{i>j} \underbrace{d_X(x_i,x_j)\, d_{ij}^{-1}(Z)}_{a_{ij} = a_{ji} \quad (a_{ij} = 0 \ \text{ for } i = j)} \left(\langle z_i, z_i\rangle + \langle z_j, z_j\rangle - 2\langle z_i, z_j\rangle\right)$$

$$= \sum_{i>j} a_{ij} \left(\langle z_i, z_i\rangle + \langle z_j, z_j\rangle - 2\langle z_i, z_j\rangle\right)$$

$$= \sum_{i>j} a_{ij} \left(\langle z_i, z_i\rangle - \langle z_i, z_j\rangle\right) + \sum_{i>j} a_{ji} \left(\langle z_j, z_j\rangle - \langle z_j, z_i\rangle\right)$$

$$= \sum_{i,j} a_{ij} \left(\langle z_i, z_i\rangle - \langle z_i, z_j\rangle\right)$$

Quadratic stress (Term 2)

$$\sum_{i>j} d_{ij}(Z) d_X(x_i, x_j) = \sum_{i,j} a_{ij}\left(\left\langle z_i, z_i\right\rangle - \left\langle z_i, z_j\right\rangle\right)$$

$$= \mathrm{tr}(BZZ^{\mathrm{T}}) = \mathrm{tr}(Z^{\mathrm{T}} BZ)$$

$$\text{where} \quad B_{ij} = \begin{cases} -a_{ij} & i \neq j \\ -\sum_{k \neq i} B_{ik} & i = j \end{cases}$$

Check:

$$(BZZ^{\mathrm{T}})_{ii} = (-\sum_{k\neq i} B_{ik})\left\langle z_i, z_i\right\rangle + \sum_{j\neq i} -a_{ij}\left\langle z_j, z_i\right\rangle = (-\sum_{k\neq i} -a_{ik})\left\langle z_i, z_i\right\rangle + \sum_{j\neq i} -a_{ij}\left\langle z_j, z_i\right\rangle$$

$$= \sum_{k\neq i} a_{ik}\left\langle z_i, z_i\right\rangle - \sum_{j\neq i} a_{ij}\left\langle z_j, z_i\right\rangle = \sum_{j\neq i} a_{ij}\left(\left\langle z_i, z_i\right\rangle - \left\langle z_j, z_i\right\rangle\right)$$

Quadratic stress (Term 2)

$$\sum_{i>j} d_{ij}(Z) d_X(x_i, x_j) = \mathrm{tr}(Z^{\mathrm{T}} BZ)$$

$$\text{where} \quad B_{ij}(Z) = \begin{cases} -d_X(x_i, x_j) d_{ij}^{-1}(Z) & i \neq j,\ d_{ij}(Z) \neq 0 \\ 0 & i \neq j,\ d_{ij}(Z) = 0 \\ -\sum_{k \neq i} B_{ik} & i = j \end{cases}$$

We make explicit the dependence of B on Z by writing $B(Z)$.

Least-squares MDS

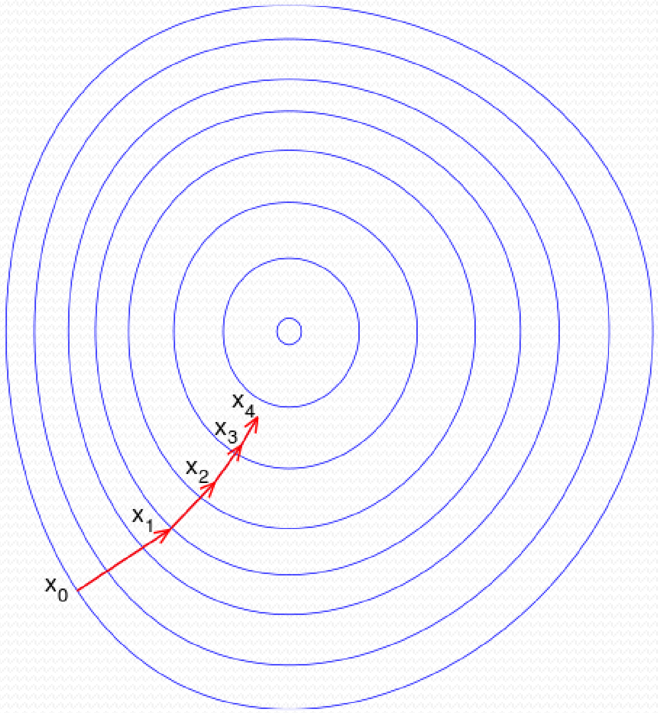
$$\sigma_2(Z) = \sum_{i>j} \left| d_X(x_i, x_j) - d_{ij}(Z) \right|^2$$
$$= \mathrm{tr}(Z^{\mathrm{T}} V Z) - 2\mathrm{tr}(Z^{\mathrm{T}} B(Z) Z) + \sum_{i>j} d_X^2(x_i, x_j)$$

Our task is to solve the unconstrained non-convex problem:

$$\min_{Z \in \mathbf{R}^{N \times m}} \sigma_2(Z)$$

We will use *gradient descent*.

Gradient descent



$$\min \ f(\mathbf{x})$$

Allows to find a *local* minimum of f.

Choose starting point $\mathbf{x}^{(0)}$

Iterate: $\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} - \alpha \nabla f(\mathbf{x}^{(t)})$

The recursive equation produces a non-increasing sequence

$$f(\mathbf{x}^{(0)}) \geq f(\mathbf{x}^{(1)}) \geq f(\mathbf{x}^{(2)}) \cdots$$

Gradient of the quadratic stress

$$\min_{Z \in \mathbf{R}^{N \times m}} \sigma_2(Z)$$

$$\nabla \sigma_2(Z) = \nabla \left(\mathrm{tr}(Z^{\mathrm{T}} V Z) - 2\mathrm{tr}(Z^{\mathrm{T}} B(Z) Z) + \sum_{i>j} d_X^2(x_i, x_j) \right)$$

$$= 2VZ - 2B(Z)Z$$

Exercise: Derive the expression given for the gradient $\nabla \sigma_2(Z)$

Gradient descent

$$\min_{Z \in \mathbf{R}^{N \times m}} \sigma_2(Z)$$

Start with a random configuration of points $Z^{(0)}$

Apply the recursive equations:

$$Z^{(t+1)} = Z^{(t)} - \alpha \nabla \sigma_2(Z^{(t)}) = Z^{(t)} - 2\alpha \left(VZ^{(t)} - B(Z^{(t)})Z^{(t)} \right)$$

Terminate when $\sigma_2(Z^{(t+1)}) / \sigma_2(Z^{(t)}) < 10^{-3}$

Multidimensional scaling

Demo Time!

Suggested reading

- *Numerical geometry of non-rigid shapes*. Bronstein, Bronstein, Kimmel. Chapters 7.1, 7.2, 7.3, 7.9