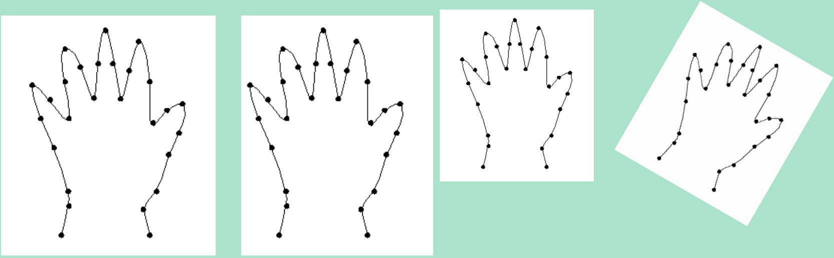
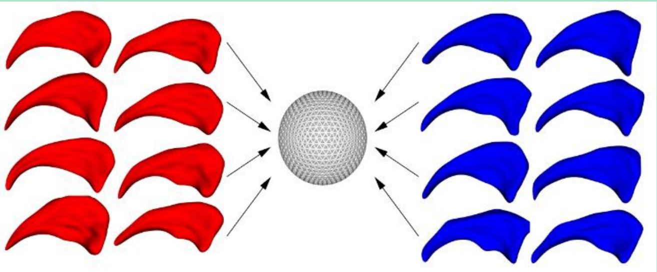
Statistics on PDMs



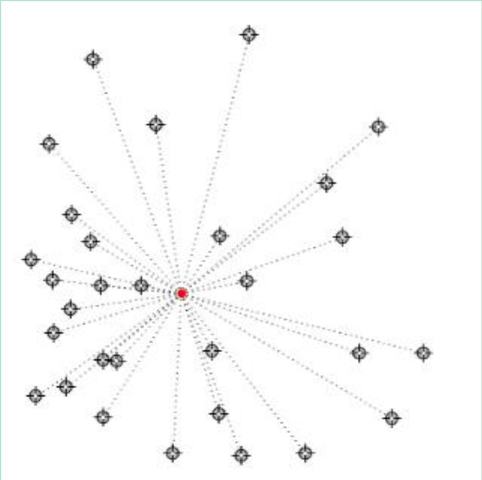
- PDM representation
- Alignment
- Principal component analysis for feature reduction
  - Forward
  - Backward
- Principal nested spheres for feature reduction
- Kendall shape space for points in 2D
- Transformation to spherical harmonics coefficients



- Find software and tutorials on slicersalt: salt.slicer.org

PDM Representation & Alignment

- PDM representation
  - $(\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_N)$
    - In 2D $(x_1, y_1, x_2, y_2, \ldots, x_N, y_N)$, a 2N-tuple



    - In 3D $(x_1, y_1, z_1, x_2, y_2, z_2, \ldots, x_N, y_N, z_N)$, a 3N-tuple
    - 3N features, understood initially in R^{3N}
- Procrustes alignment: Least sum of squares fit over coordinates
  - Centering: by subtracting center of mass
  - Scaling (after centering)
  - Rotation so axes of best fitting ellipsoid are in the cardinal directions

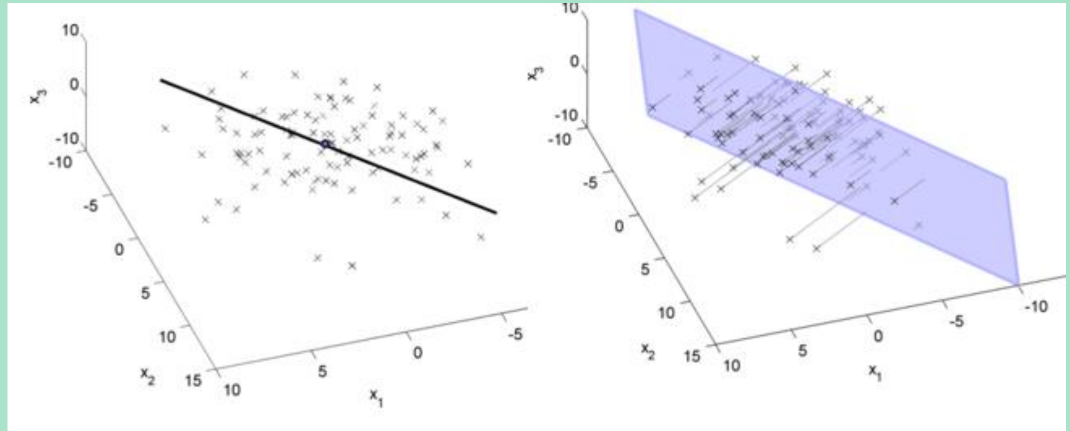
Alignment of PDMs

- PDM representation
  - $(\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_N)$
    - So in 3D $(x_1, y_1, z_1, x_2, y_2, z_2, \ldots, x_N, y_N, z_N)$, a 3N-tuple
    - 3N features, understood initially in R^{3N}
- Procrustes alignment
  - Centering: subtract center of mass: $\underline{x} - \bar{\underline{x}}$
    - You have removed 3 degrees of freedom, so in R^{3N-3}
  - Scaling (after centering): divide by $\Sigma_i(x_i^2+y_i^2+z_i^2)^{½}$
    - You have removed 1 additional degree of freedom, has dimension 3N-4
    - But now $\Sigma_i(x_i^2+y_i^2+z_i^2) = 1$, so on the unit sphere S^{3N-4}
  - Rotation: Rotate to eigenvectors of the 3×3 2nd moment matrix, whose jkth entry is Σ_i of the jth among (x,y,z) times the kth among (x,y,z)
    - These are the axes of the best fitting ellipsoid
    - Removed 3 more degrees of freedom so result on S^{3N-7}

Principal Component Analysis

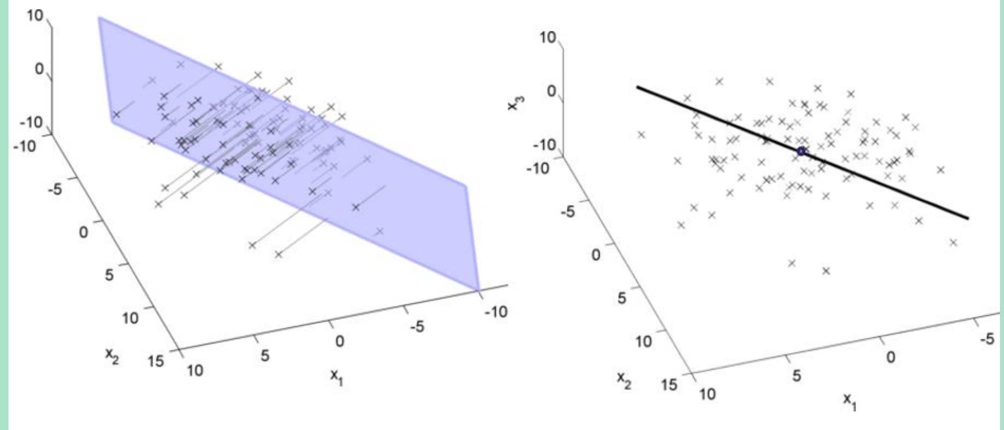
- For feature reduction from dimension N to much smaller
- Rotate feature coordinates such that features in the new (principal) coordinates are uncorrelated and ordered to have the first m features capture the data best, for all m
  - "Best" means capture most of the population's variation
  - So later principal features can be ignored, as representing mostly measurement noise
- Principal directions computed as eigenvectors of the estimated covariance matrix of the raw features
  - Via AA^T, where rows of A are the individual data tuples with their mean subtracted
- Principal variances σ_k^2 computed as eigenvalues of the estimated covariance matrix
- Ordered in decreasing order of principal variances
- If # of training samples $n < N$, $\sigma_k^2 = 0$ for $k > n-1$
  - Because an n-1-dimensional hyperplane matches all n points

Forward PCA

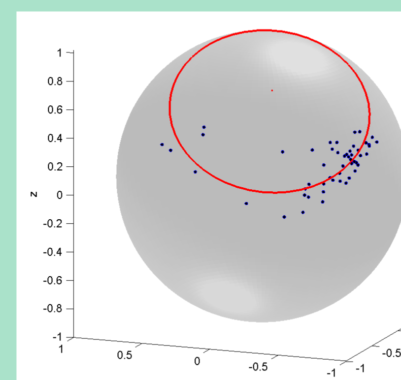


- By increasing order of dimension of flat (Euclidean) representation
  - Dimension 0: point; dimension 1: line; dimension 2: plane, …
- Best summarizing point in feature space: arg $\min_{\underline{x}} \Sigma_\iota \, d(\underline{x}_i, \underline{x})^2$, the Fréchet mean (the ordinary average if Euclidean distance is used)
- Best summarizing line in feature space: arg $\min_{line} \Sigma_\iota \, d(\underline{x}_i, \text{line})^2$
  - Passes through the best summarizing point
  - The unit vector in the line is the 1st principal component
- Best summarizing plane in feature space: arg $\min_{plane} \Sigma_\iota \, d(\underline{x}_i, \text{plane})^2$
  - Passes through the best summarizing line
  - The unit vector orthogonal to the 1st princ. comp't is the 2nd principal component
- Best summarizing flat 3D space within feature space
  - Passes through the best 2D representation
  - The unit vector orthogonal to the best summarizing plane is the 2nd principal component
- …

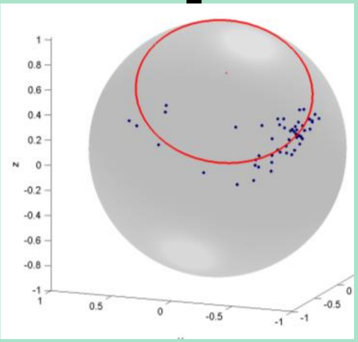
Backward PCA

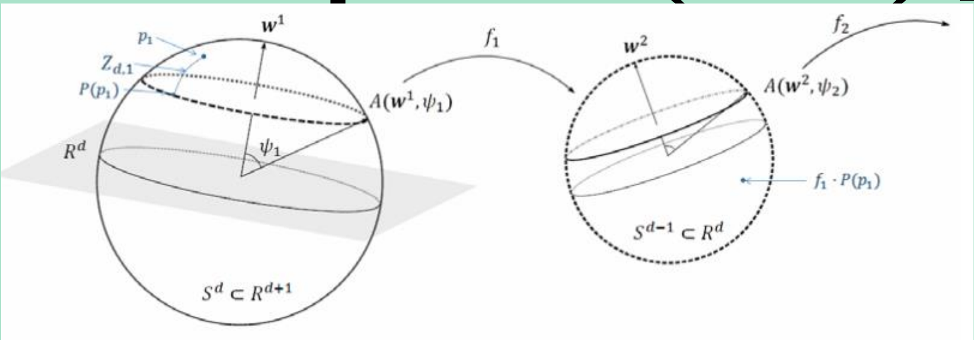


- By decreasing order of dimension of flat representation
  - Dimension N-1; dimension N-2,; …; dimension 1: point; dimension 0: point
  - When going from dimension k to dimension k-1, **project the data** from the dimension k space along its geodesic to the dimension k-1 space, and **the projection distance for each data point becomes the kth principal feature**
  - Decreasing dimension from d to d-1 involves least squares fitting of d-1 dimensional hyperplane to data on d-dimensional hyperplance
- For Euclidean feature space,
  gives the same result as forward PCA
- But for feature spaces that are curved manifolds, it is better than forward PCA because the subspaces stay within the data
  - It fart fails to do that for data near a great circle on a sphere



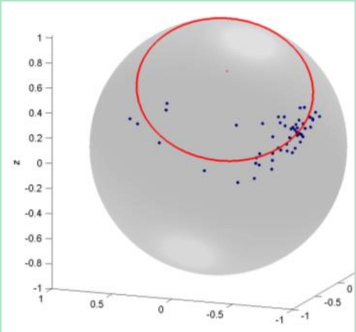
Principal Nested Spheres (PNS) [Jung]

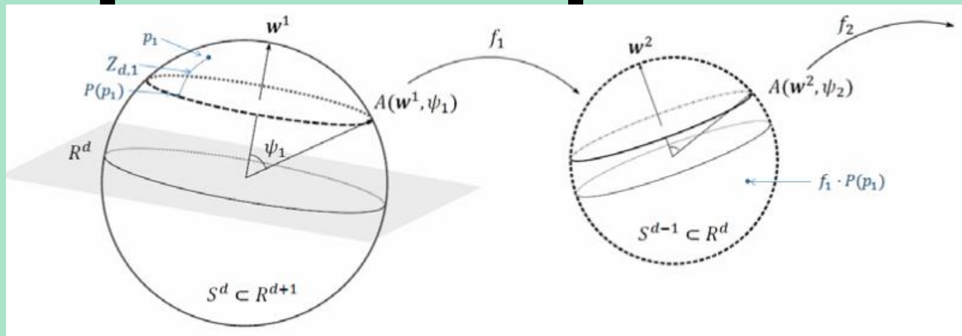




- Is backwards PCA on spheres transiting into sub-spheres
  - Dimension N-1; dimension N-2,; …; dimension 1: line; dimension 0: point (**backwards mean**)
  - When going from dimension k to dimension k-1, project the data from the dimension k space along its geodesic to the dimension k-1 space, and **the projection distance for each data point becomes the kth principal feature**
- Examples: 1) direction feature: on S^2, 2) on S^m, 3) on polysphere $(S^2)^N$
  - On S^2 produces 2 features; on S^m produces m features
- For dimension > # of training samples n, can fit an n-1 dimensional hyperplane exactly, then start decreasing dim'ns

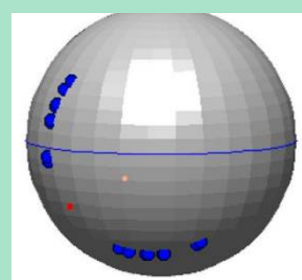
PNS Example: on 2-sphere



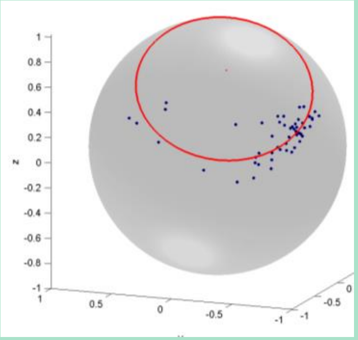


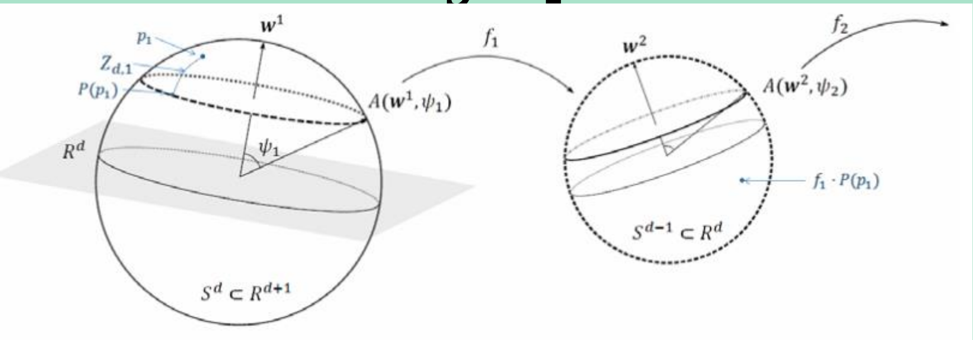
  - When going from dimension k to dimension k-1, project the data from the dimension k space along its geodesic to the dimension k-1 space, and **the projection distance for each data point becomes the kth principal feature**
- When feature is a direction, feature space is S^2
  - Find best fitting subsphere of dimension 2-1 = 1, a circle
    - On the containing sphere, and data is projected onto the circle
    - Derived feature 2 for a data point is arc length (angle) from the point to the circle
  - Find best fitting subsphere of dimension 1 -1 = 0, a point: **the backward mean**
    - On the containing circle, and data is projected onto the point
    - Derived feature 1 is also an angle

Backwards mean (light) is better than Fréchet mean (red)



PNS on Polysphere





- For tuple of N directions, data is on polysphere $(S^2)^N$; apply PNS 2-sphere by 2-sphere
  - Obtaining 2N features, 2 per 2-sphere
  - There is still correlation among the feature-pairs, so a PCA on the 2N derived feaures is needed

Principal Subspaces for Other Manifolds

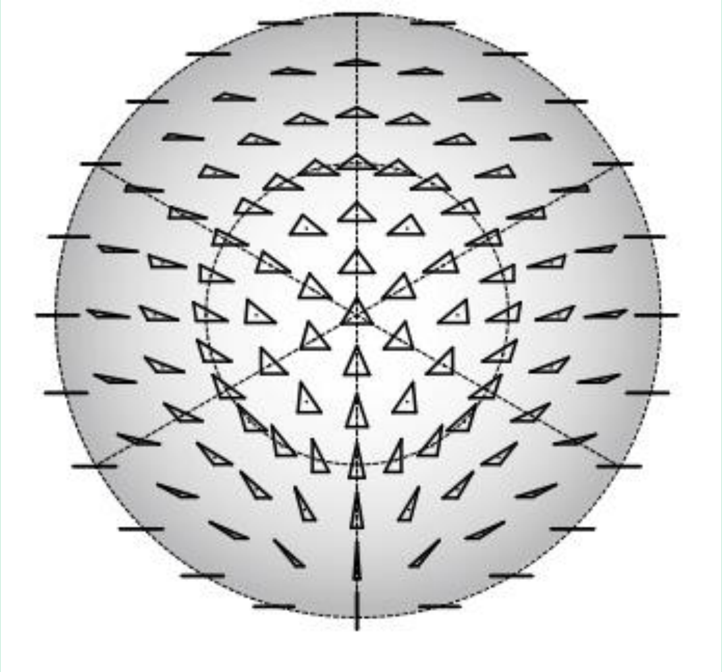
- What succession of subspaces?
  - In particular for polyspheres
- Possibility that all subspaces and the associated derived features are computed simultaneously rather than successively [X Pennec]: Barycentric analysis

Kendall Shape Space

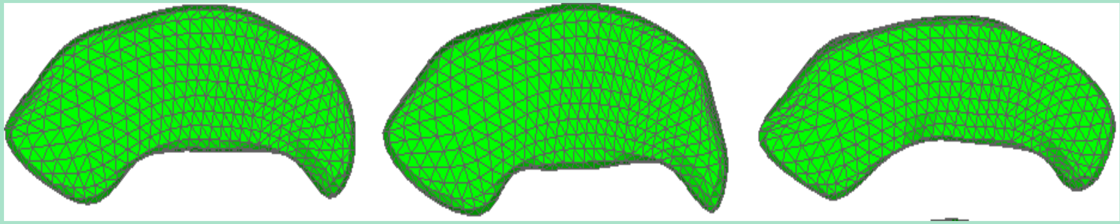
- For k points in Euclidean 2-space but understood as complex numbers z_i
  - There is extension to 3D, but complicated
- Remove translation
- For k points, understood as the complex, projective space CP^{k-2}
- In complex numbers space, a scaling and rotation of $z=re^{i\theta}$ is accomplished by a multiplication by $w= \alpha e^{i\phi}$
  - α gives the scale factor, ϕ gives the rotation angle
- Thus CP^{k-2} is {wz| modulo w}
- Also, each PDM in population is produced by removal of translation, scale, and rotation, so is on the unit sphere, so statistics are derived using distances on the unit sphere
- Ref: [Dryden & Mardia, *Statistical Shape Analysis*, either 1998 edition (on reserve) or 2016 edition (on line)]

Statistics on Kendall Shape Space

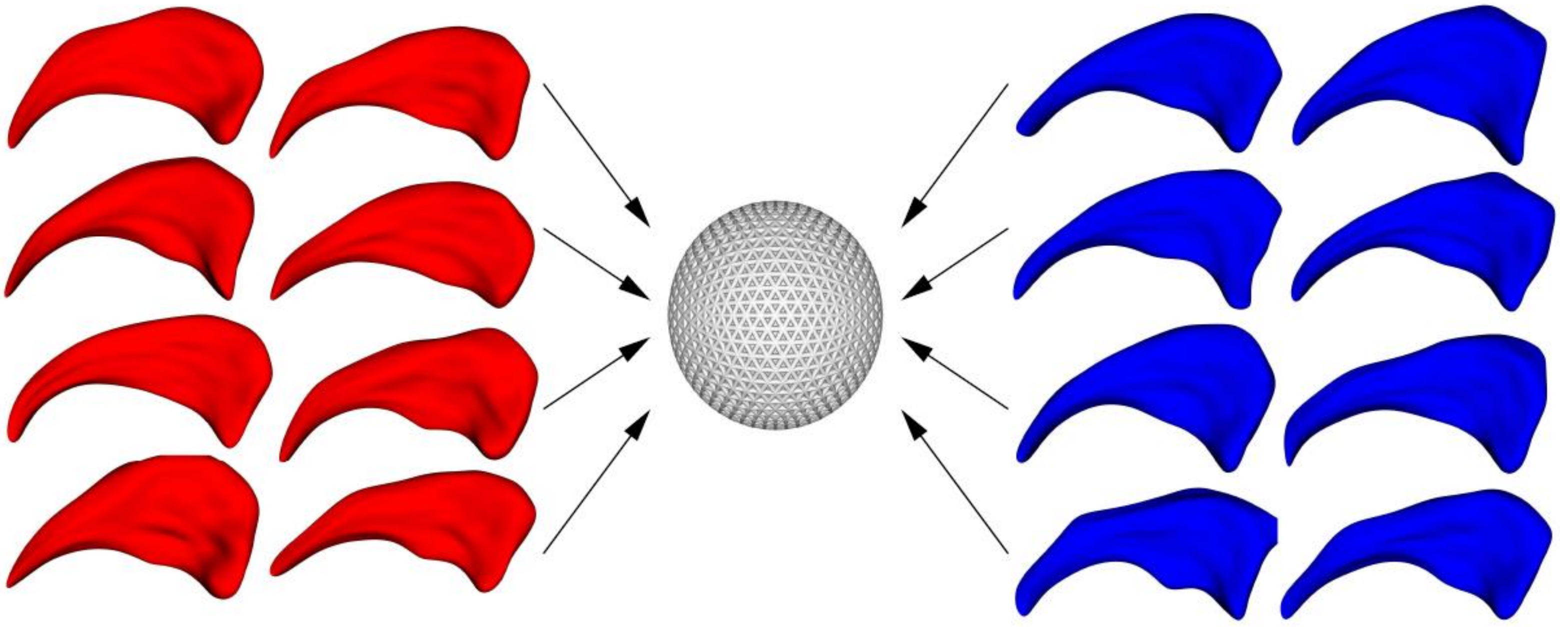
- Distances are between inter-3-point triangles
  - After the normalizations, and with points correspondence
  - Procrustes distances among corresponding points
  - For only 3 points for n=2 or 3, 3n-7=2, i.e, triangles live on S^2



- Inter-triangle tuple distances are formed from Riemannian distances among the corresponding triangles
  - Various choices for which triangles to use and how to derive the combination



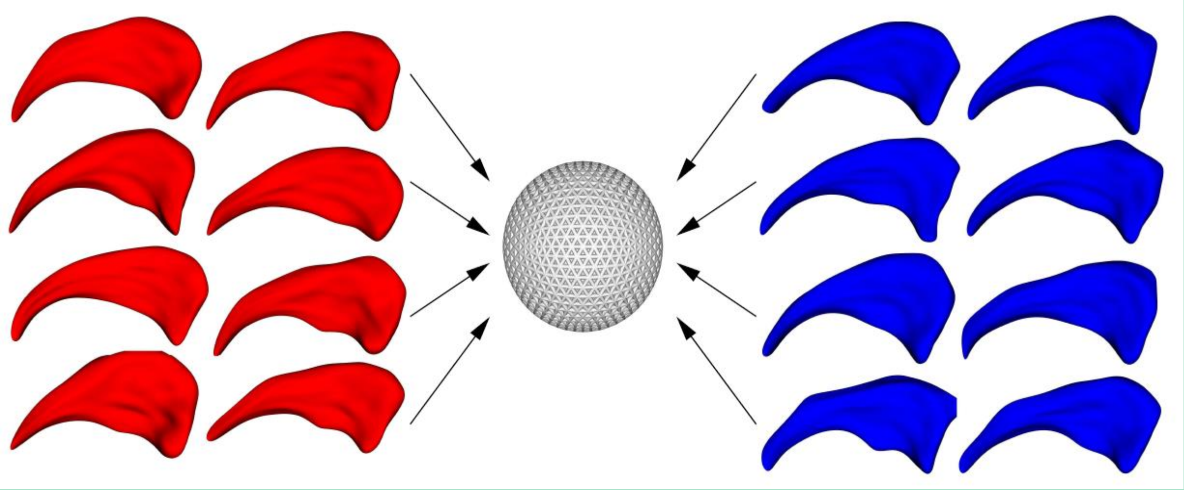
Representations on the Sphere



Objects described by basis functions on the sphere.
Challenge: How to get the point coordinates onto the object in the first place.

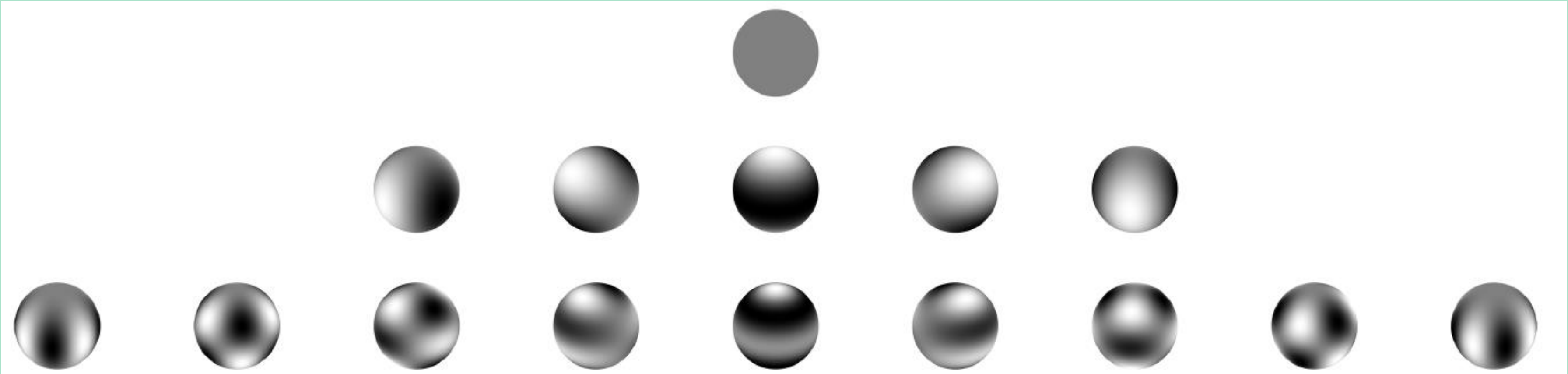
Statistics on PDMs Transformed into Spherical Harmonics Coefficents

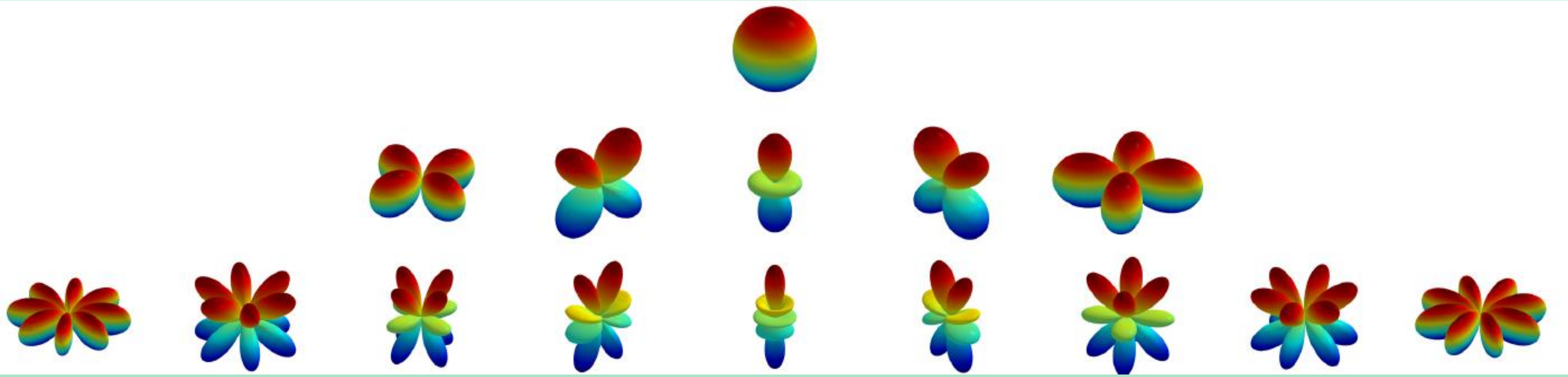
- Each object mapped from sphere: $\underline{x}(\theta,\phi)=\Sigma_i \underline{b}_i \; \psi^i(\theta,\phi)$
  - Discretized with equal area spherical triangles
  - Can do Euclidean statistics of the $\underline{b}$ values over a population
  - $\underline{b}$ values are determined globally



- Object features: coefficients of basis functions on the sphere
  - Basis functions organized by frequency in latitude and longitude
  - From $\underline{x}(\theta,\phi)$, coefficients easily obtained by dot product w/ basis
  - For any (θ,ϕ), $\underline{x}(\theta,\phi)$ (e.g., mean) can be computed from coefficients
    - **Correspondence via (θ,ϕ)**, but empirically not always adequate

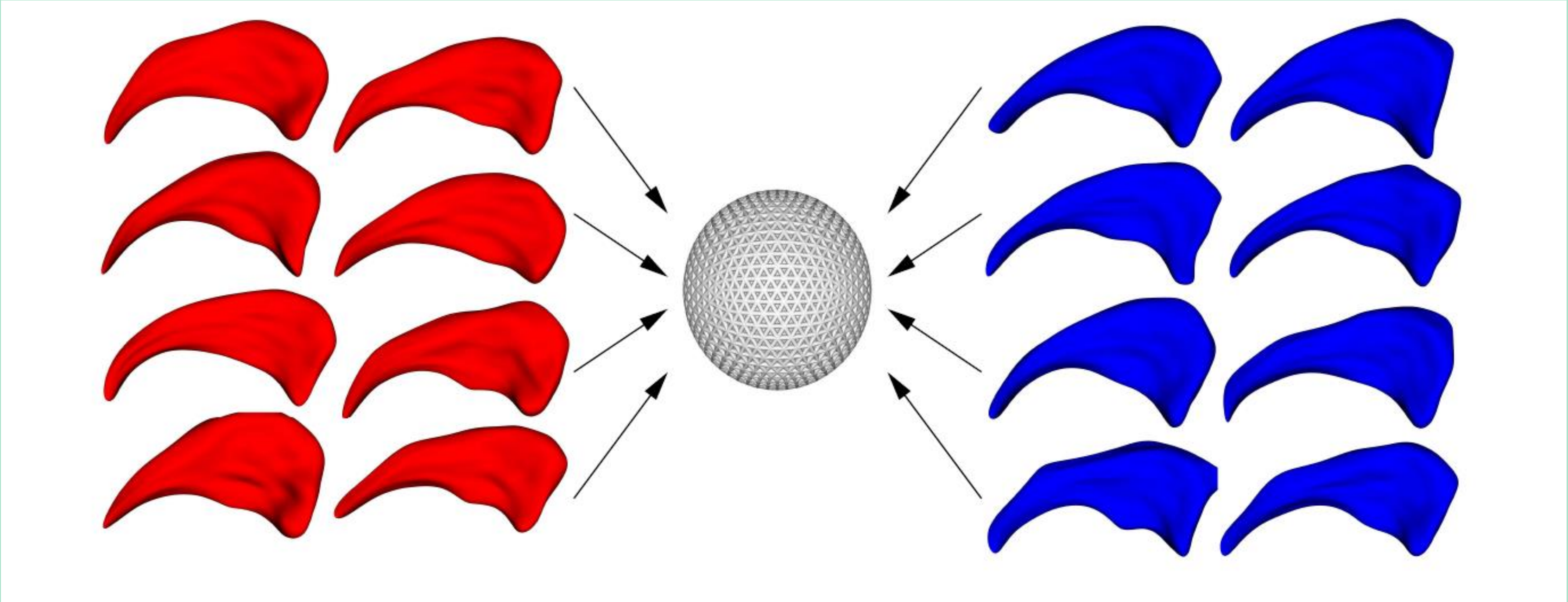
Spherical Harmonics Basis Functions



- 

How to get the point coordinates on the object onto the sphere

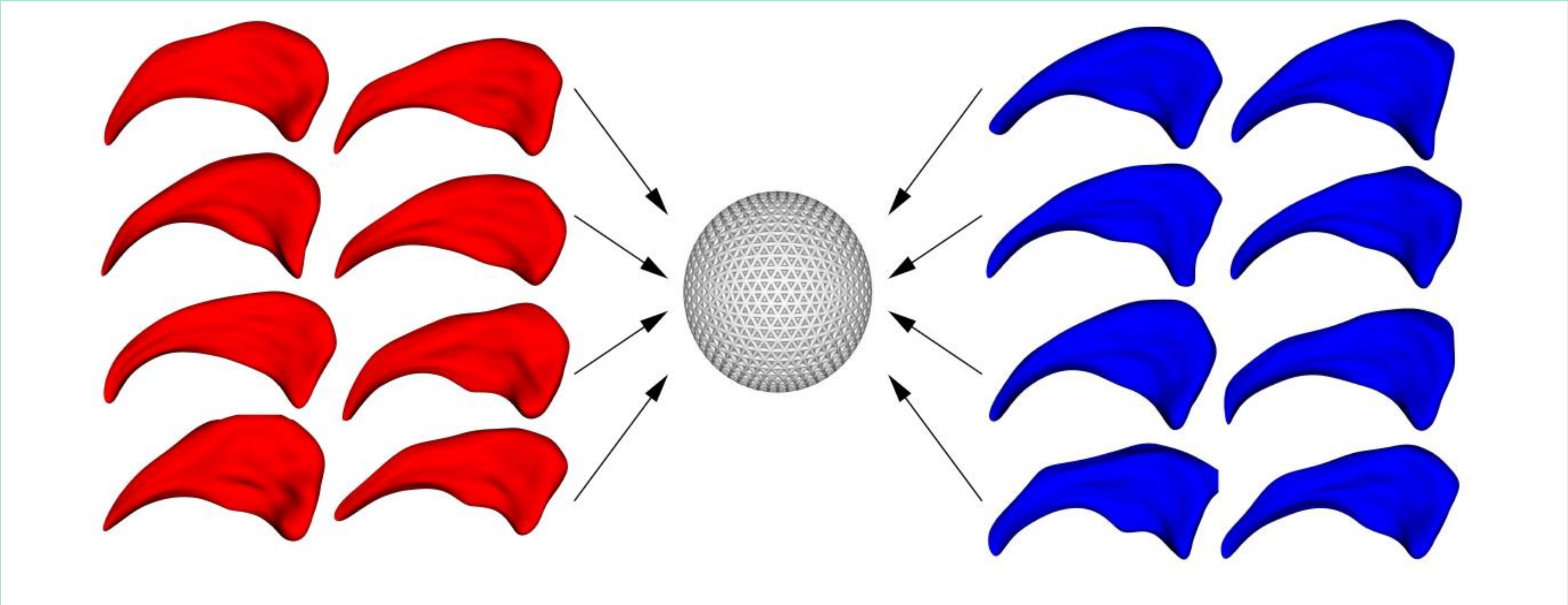
- Equal area mapping [Brechbuehler]
  - Or alternative of conformal mapping, which is angle preserving
- North pole and Greenwich meridian via best fitting ellipsoid
  - Might need straightening as a preprocessing



- Possible use of s-reps implied spacing

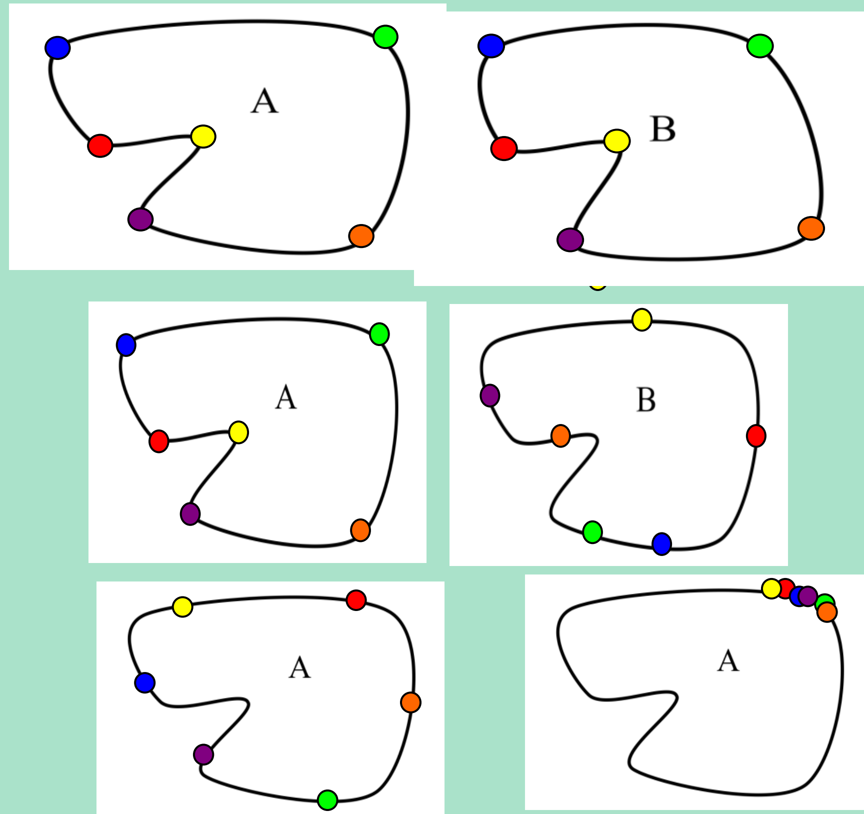
Correspondence for PDMs

- Get initial correspondence
- Entropy optimization [ShapeWorks]



Correspondence

- Approaches
  - Via entropy: produce tightest ensemble p($\underline{\boldsymbol{x}}$)
    - Possibly also including C, S as features [Oguz]
  - Registration
    - Via landmarks
      - thin plate splines
      - diffeo guaranteeing methods
    - Via richer geometry, such as skeletal



Tight distr'n (low ensemble entropy)

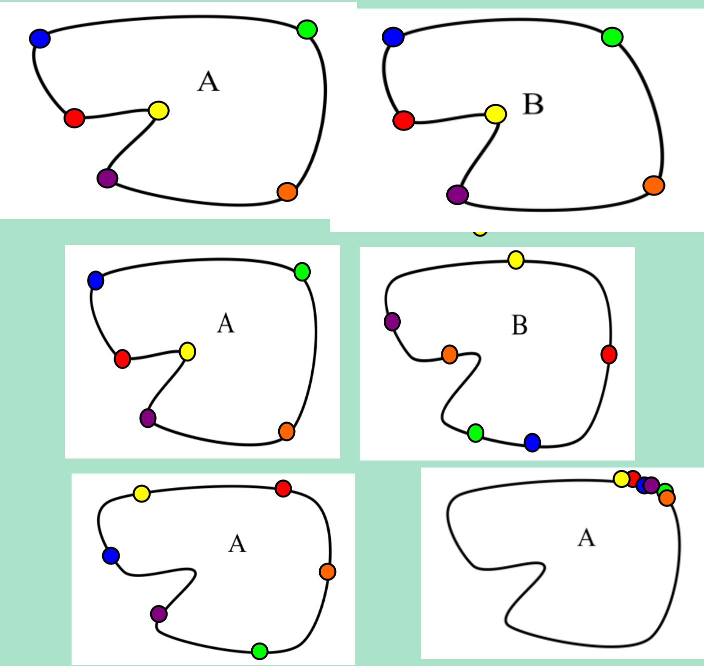
Non-tight distr'n (high ensemble entropy)

High surface entropy

Low surface entropy (uniformity)

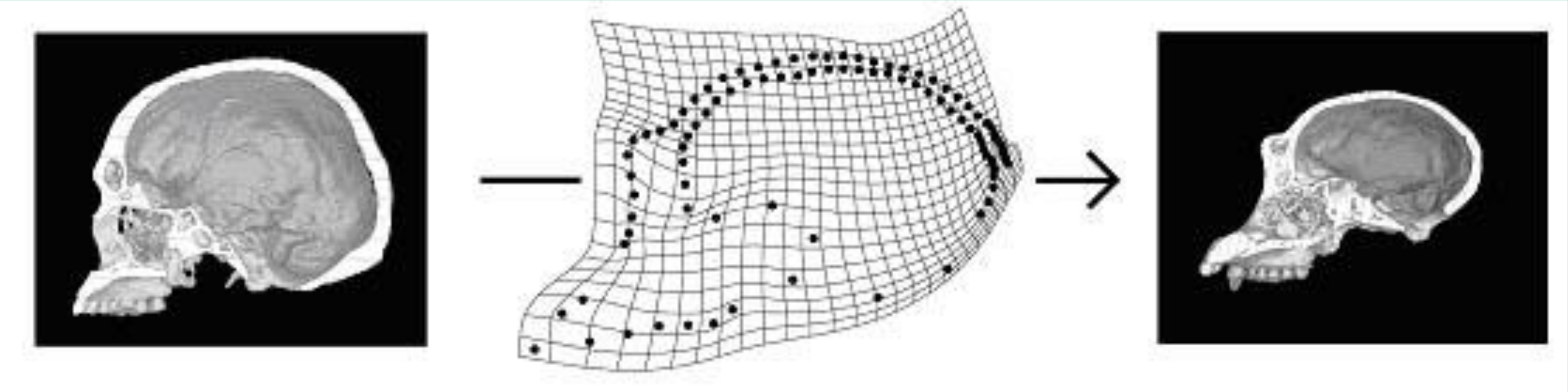
Correspondence via Entropy of PDMs

- Shapeworks [Cates, Whitaker]
  - Ensemble entropy H(ensemble) should be low (p(**x**) tight)
  - Entropy H(point positions along boundary for each case) should be high (uniformly distributed)
  - So $\min_{\underline{\mathbf{x}}}$ [$H_{\text{training cases}}$(geometry) – $\Sigma_{\text{training cases}}$ H(points on training case)]
  - Entropy via PCA: H(nD Gaussian) = (n/2) [1+ ln(2π) + avg ln λ]
  - Optimize by successively doubling number of points
    - Slow and often finds local optimum
- Better stats come from ridiculously inefficient entropies from s-rep [Tu, Vicory et al.]



Correspondence via Landmark Registration

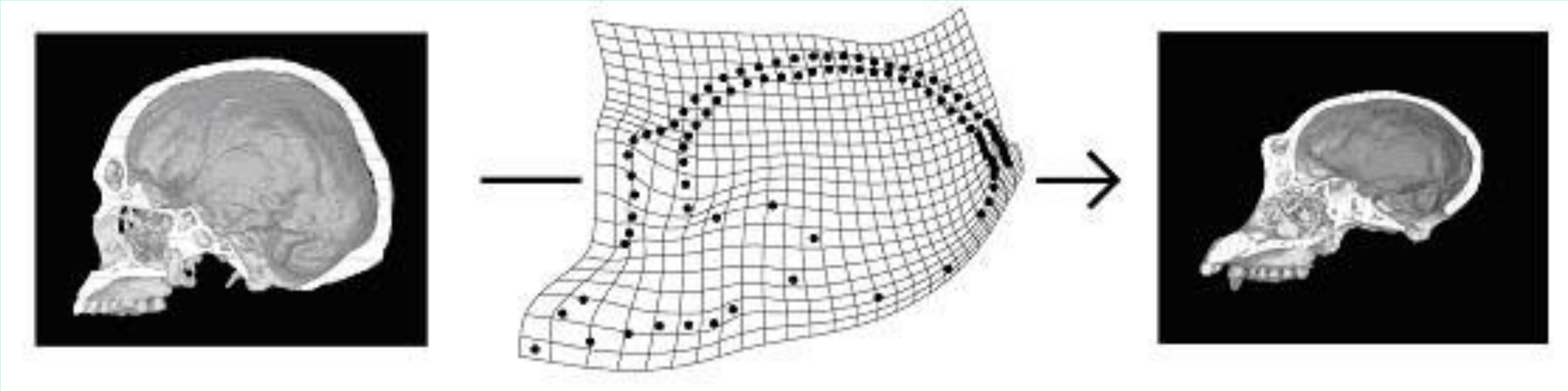
- Typically preceded by affine registration
- Means of interpolation of corresponding landmarks $(\underline{x}_i, \underline{x}'_i)$ to continuous deformation $\underline{x}'(\underline{x})$



  - Thin plate splines
  - Methods guaranteeing diffeomorphisms via LDDMM
    - Joshi
    - Deformetrica
    - Symmetry guaranteeing methods

Thin Plate Splines for Landmark Based Deformation

- Optimum (perfect) data match, with geometric typicality (smoothness) analytically minimum
  - Compute continuous deformation $\underline{x}'(\underline{x})$ from landmarks



Warping a human skull into a chimpanzee skull.

  - Fast: based on a solution to linear equations
  - Typically preceded by optimum affine transformation

Thin Plate Splines Method

- Elastic warp in each variable
  - $\underline{\mathbf{x}}'(\underline{\mathbf{x}})=\underline{\mathbf{c}}+\mathbf{A}\underline{\mathbf{x}}+\Sigma_{\mathbf{j}}\, \underline{\mathbf{w}}_{\mathbf{j}}\mathbf{U}(|\underline{\mathbf{x}}-\underline{\mathbf{x}}^{\mathbf{j}}|)$
  - Basis functions $U(|\underline{x}-\underline{x}^j|)$ depend on moving image's landmarks $\underline{\mathbf{x}}^{\mathbf{j}}$
    - Radial bases: $U(d) = d^2 \log d$ for 2D, d^3 for 3D
- Solve linearly for c, A, $\{\underline{\mathbf{w}}_{\mathbf{j}}\}$ based on $\{\underline{\Delta\mathbf{x}}^{\mathbf{j}}\}$
- Minimizing Frobenius norm: $\int^{\infty \text{ space}} \Sigma_{\text{all}}$ 2nd partial derivatives^2, so smooth
  - 27 terms for 3D: 9 for $\Delta x(x,y,z)$, 9 for $\Delta y(x,y,z)$, 9 for $\Delta z(x,y,z)$
- Not necessarily diffeomorphic; may produce folding
  - Normally OK if displacements << inter-landmark spacing
- Not symmetric, not affine invariant
- Due to Bookstein: Ref: [Dryden & Mardia, *Statistical Shape Analysis*, either 1998 edition or 2016 edition]

Large Deformation Diffeometric Metric Mapping (LDDMM) Methods

- Consider the shape space of diffeomorphisms
- Let metric on that space measure spatial smoothness within a velocity image
- We want the shortest geodesic from Identity mapping to the diffeomorphism that maps the corresponding points onto each other
- Typically requires iterative optimization
- Implementations
  - Joshi (see next slide)
  - Deformetrica (see later in course)
    - Can also use corresponding space curves

Diffeomorphic Landmark Matching [Joshi]

Flowing images into each other. Mapping function $h(\mathbf{x}) = \phi(\mathbf{x}, 1)$ given through the ODE

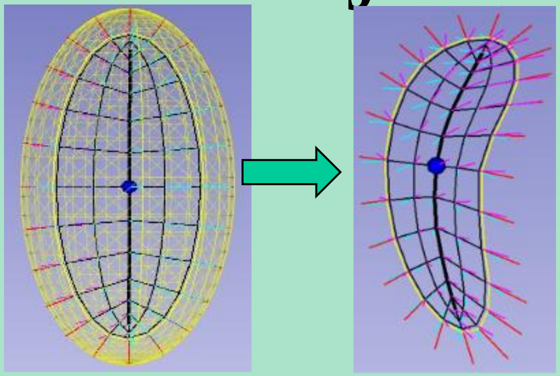
$$\frac{d\phi(\mathbf{x}, t)}{dt} = v(\phi(\mathbf{x}, t)), \quad t \in [0, 1], \quad \phi(\mathbf{x}, 0) = \mathbf{x}.$$

Minimize smoothness cost subj. to landmark constraints $(h(\mathbf{x}_n) = \mathbf{y}_n)$

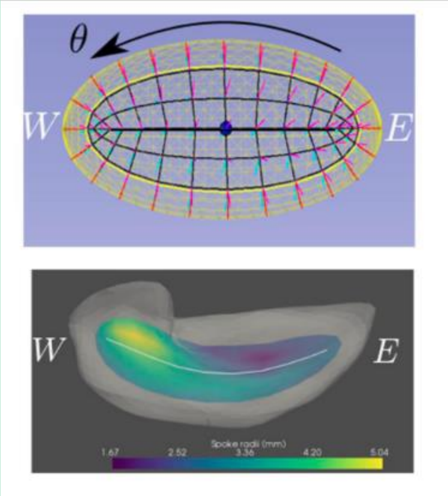
$$\hat{v}(\cdot) = \underset{v(\cdot)}{\operatorname{argmin}} \int_0^1 \int_\Omega \|Lv(\mathbf{x}, t)\|^2 \, d\mathbf{x} \; dt.$$

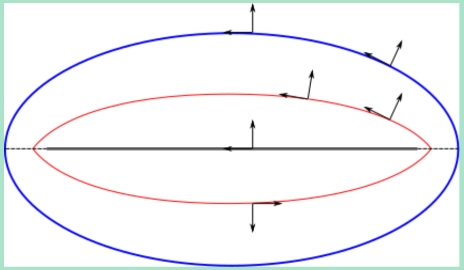
This is guaranteed to give a diffeomorphic h for suitable L (for example $L = I(-\nabla^2 + c)$ works). in 2D

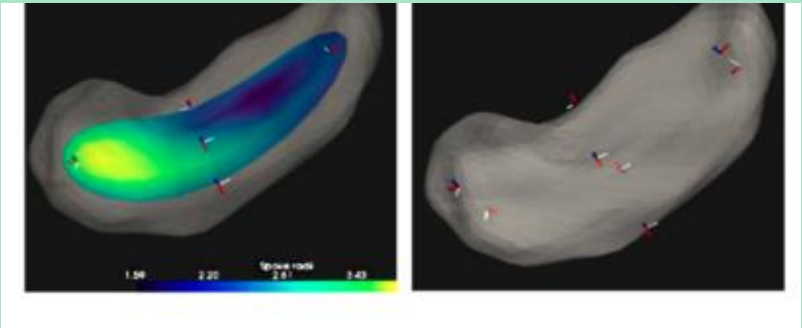
Correspondence via Skeletal Mapping from Ellipsoid to Object



- Mapping via diffeomorphism such that
  - Vertices map onto vertices
  - Crests map onto crests
  - Straight spokes map onto straight spokes
- Sampled spoke points (from skeleton to boundary) map onto each other determine diffeomorphism
- Defines a fitted frame at every sampled spoke point







- [Pizer, *Skeletons, Object Shape Statistics*, Frontiers in Computer Science, 2023, on google drive for Pizer, Comp 790-6, will be assigned