

Multidimensional scaling as a tool for smoothing over complex spatial regions

David Lawrence Miller

Mathematical Sciences
University of Bath

January 18, 2011

Outline

Smoothing over complex regions

- The problem

- Smoothing with splines

- Solutions

Multidimensional Scaling

- Details

- Finding the within-area distances

- Simulation Results

Conclusions

Outline

Smoothing over complex regions

- The problem

- Smoothing with splines

- Solutions

Multidimensional Scaling

- Details

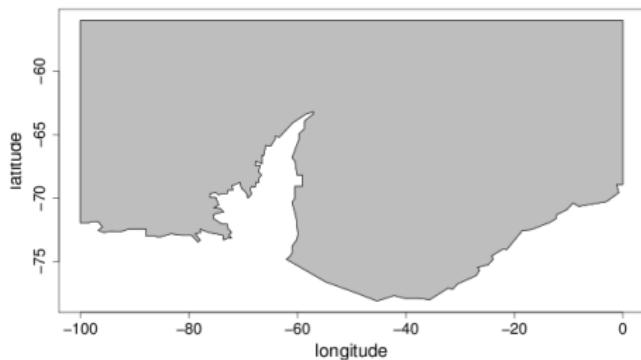
- Finding the within-area distances

- Simulation Results

Conclusions

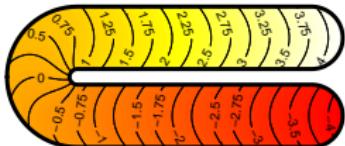
Smoothing in 2 dimensions

- ▶ Have some geographical region and wish to find out something about the biological population in it.
- ▶ Response is eg. animal distribution, wish to predict based on (x, y) and other covariates eg. habitat, size, sex, etc.
- ▶ This problem is relatively easy if the domain is simple.

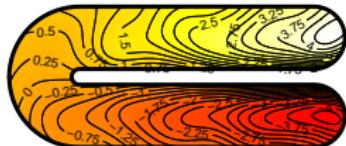


Leakage (or, “whales don’t live in glaciers”)

- ▶ Smoothing of complex domains makes this a lot more difficult.
- ▶ Problem of leakage.
- ▶ Wrong metric.
- ▶ Euclidean distance doesn’t always make sense.



(modified) Ramsay test function



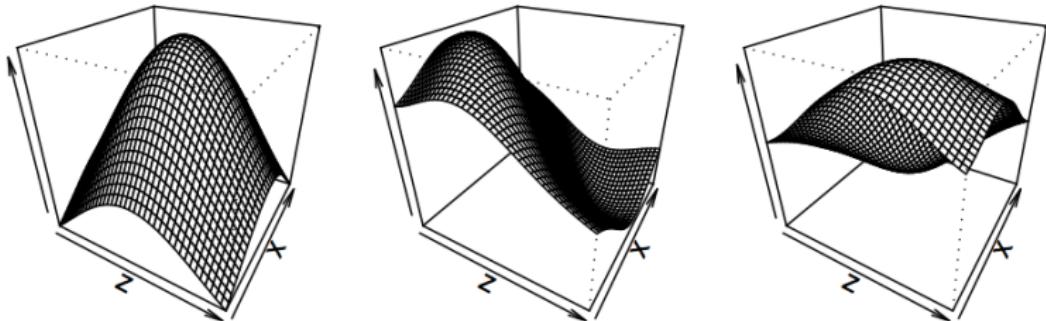
Thin plate spline fit

Smoothing with penalties

- ▶ Do smoothing with splines in an additive model framework.
- ▶ Take some linear combination of (known) basis functions.
- ▶ Penalize based on integral of second squared derivative:

$$\int \int_{\Omega} \left(\left(\frac{\partial^2 f(x, y)}{\partial x^2} \right)^2 + \left(\frac{\partial^2 f(x, y)}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 f(x, y)}{\partial y^2} \right)^2 \right) dx dy.$$

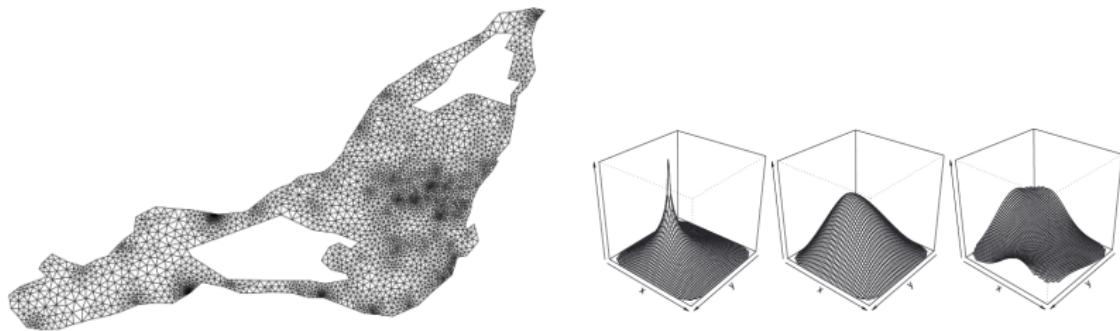
- ▶ Here using thin plate regression splines (eg Wood (2003)).



Proposed solutions to leakage problems

Two categories: PDE boundary condition based, within-area distance based.

- ▶ PDEs:
 - ▶ FELSPINE (Ramsay, 2002).
 - ▶ Soap film smoothers (Wood *et al*, 2008).
- ▶ Within-area distance
 - ▶ Geodesic Low-Rank Thin Plate Splines (GLTPS) (Wang and Ranalli, 2007).

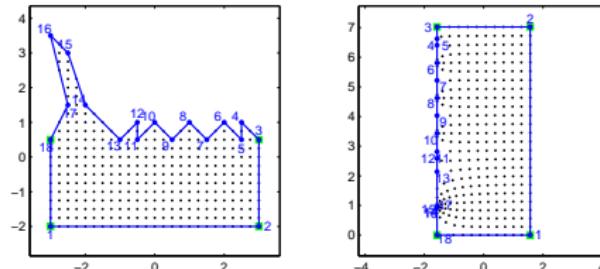


“The Third Way”

- ▶ Domain morphing - think of domain as silly putty.
- ▶ Takes into account within-area distance.
- ▶ Potentially less computationally intensive.
- ▶ Eilers (2006) proposes Schwarz-Christoffel transform.
- ▶ Gives a known domain - easier to smooth over.

However:

- ▶ Crowding - numerical singularities.
- ▶ Not clear what this does to the smoothness penalty - artefacts, penalty issues - but NOT GOOD!



Outline

Smoothing over complex regions

The problem

Smoothing with splines

Solutions

Multidimensional Scaling

Details

Finding the within-area distances

Simulation Results

Conclusions

Multidimensional scaling and within-area distances

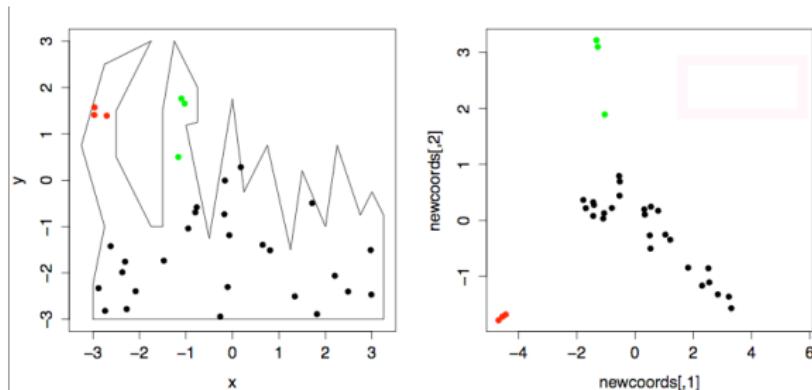
- ▶ Idea: use MDS to arrange points in the domain according to their distance within the domain.

Scheme:

- ▶ First need to find the within-area distances.
- ▶ Perform MDS on the matrix of within-area distances.
- ▶ Find the new configuration of the points.
- ▶ Smooth over the new points.

Multidimensional scaling refresher

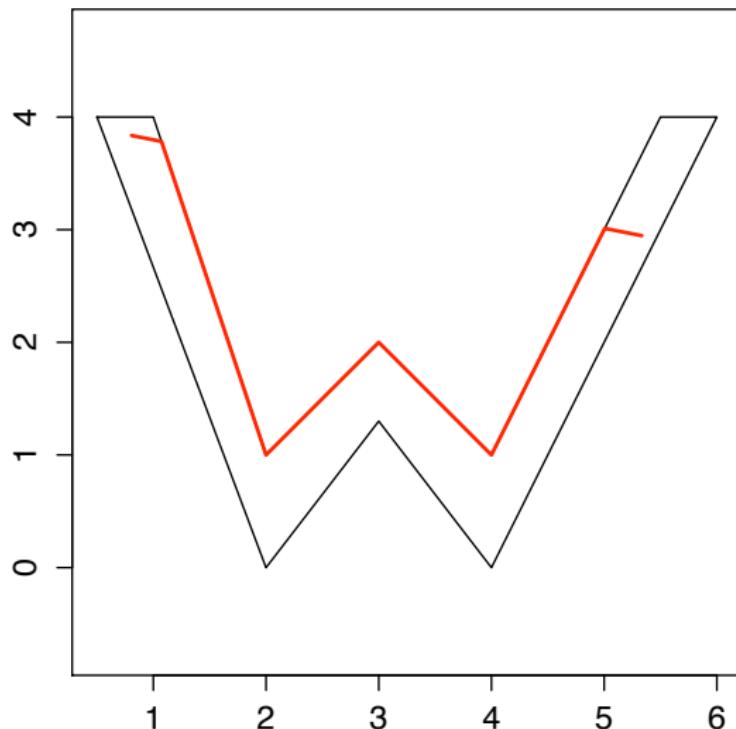
- ▶ Use double centred matrix of between point distances, D , using the eigen-decomposition of DD^T we can find new points.
- ▶ Finds a configuration of points such that Euclidean distance between points in new arrangement is approximately the same as distance in the domain.
- ▶ New points can be added into the MDS configuration via Gower's interpolation (Gower, 1968)



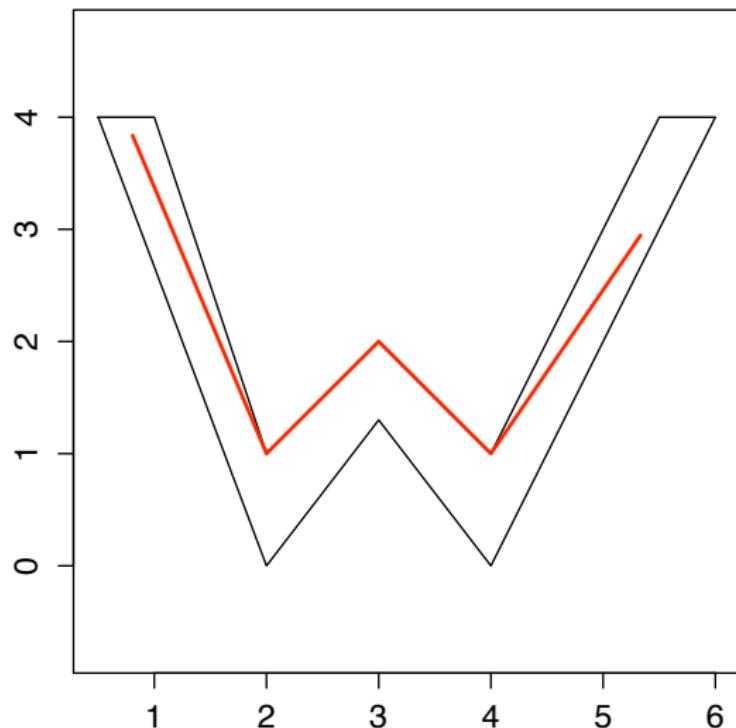
A “new” algorithm for finding within-area distances

So, how do we find the distances to put in D?

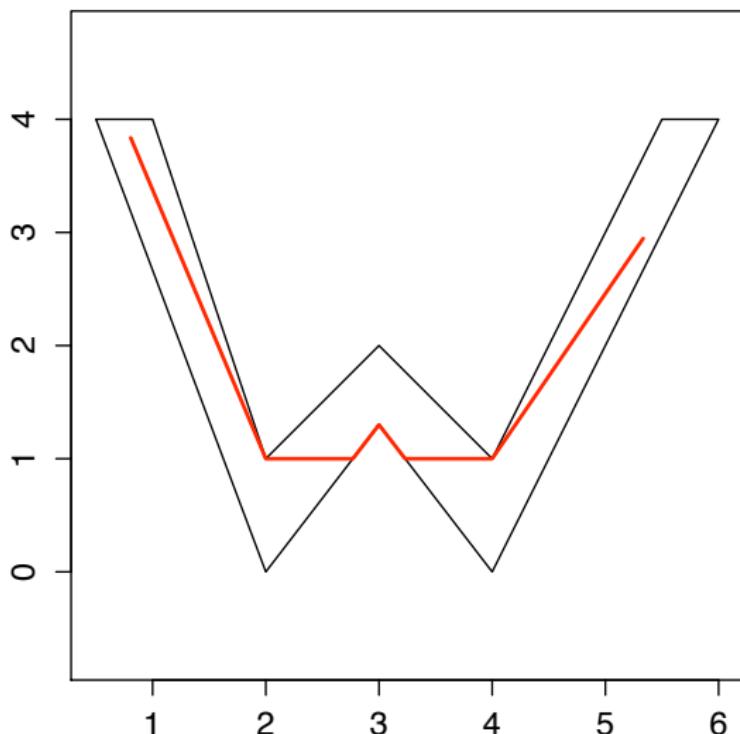
Finding the path - 1



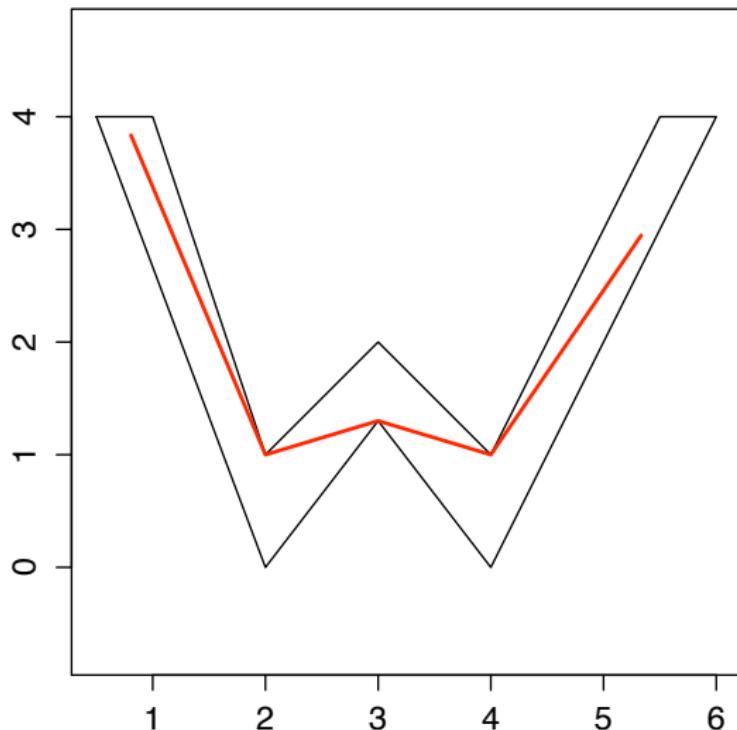
Finding the path - 2



Finding the path - 3



Finding the path - 4

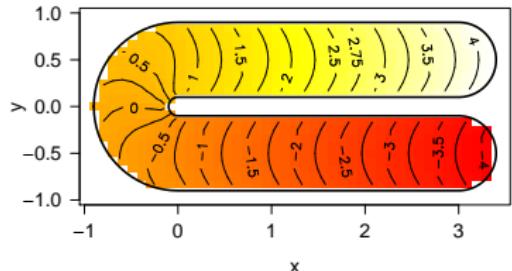


Simulations

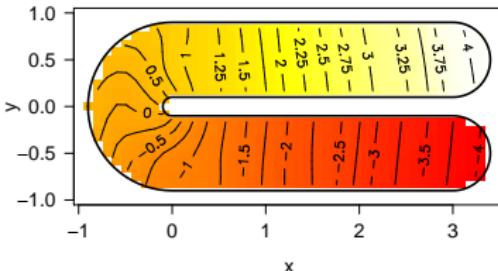
But how does it perform?

Ramsay simulations

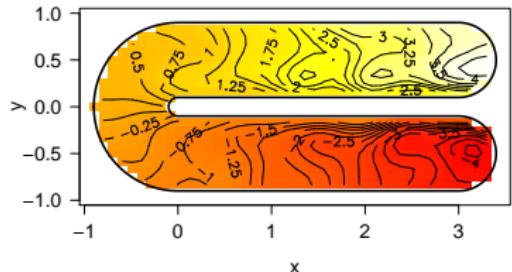
truth



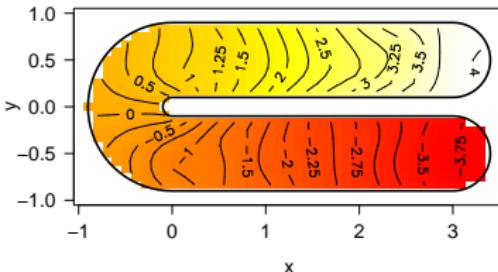
MDS



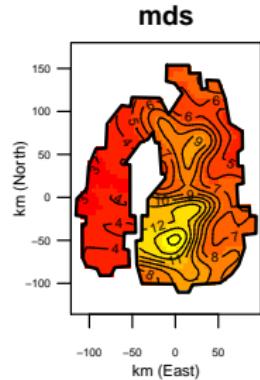
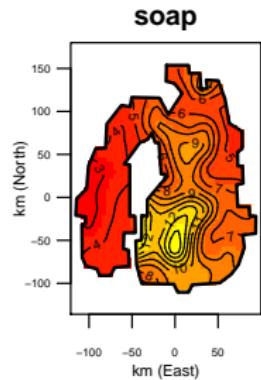
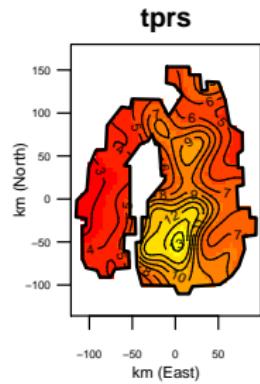
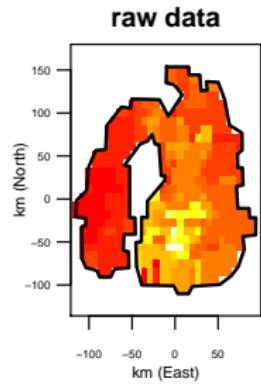
tprs



soap

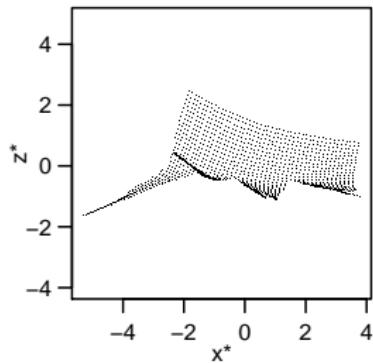
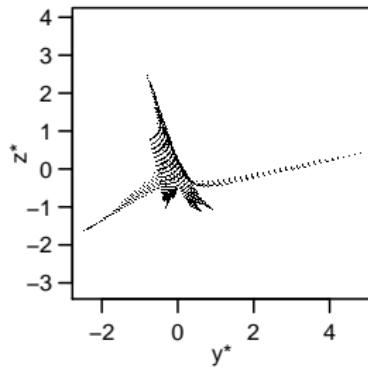
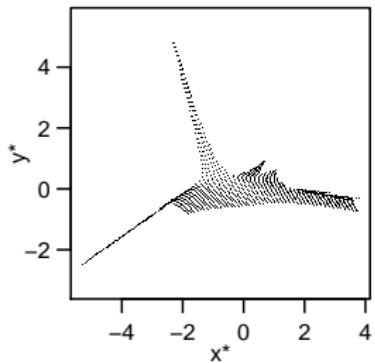


The Aral sea



So, it looks good, but...

- ▶ Computationally costly to find the within-area distances.
(Esp. prediction.)
- ▶ Comparative in MSE terms to the soap film smoother.
- ▶ Artefacts.
- ▶ Taking a m -D projection of n -D space.



Can we get around the squashing?

- ▶ For anisotropic smoothing, Wood (2000) suggests:

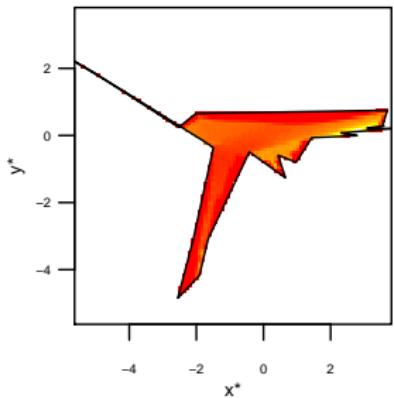
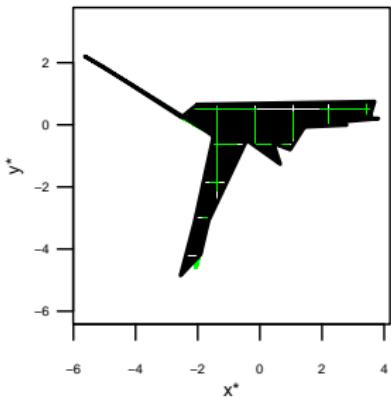
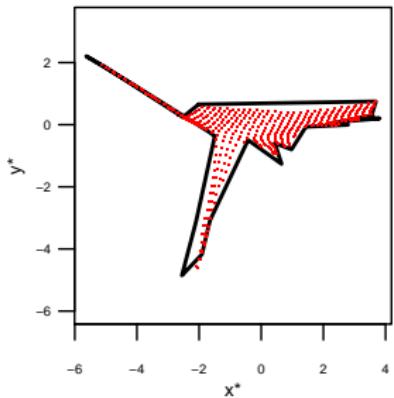
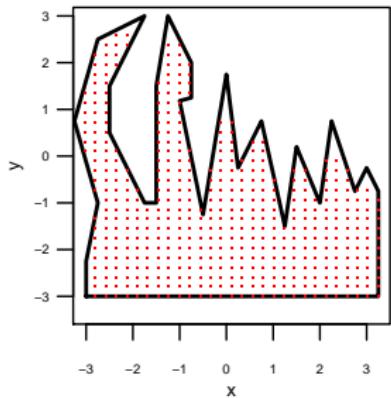
$$\int \int_{\Omega} \left(\frac{\partial^2 f(x, y)}{\partial x^2} \right)^2 + k \left(\frac{\partial^2 f(x, y)}{\partial x \partial y} \right)^2 + k^3 \left(\frac{\partial^2 f(x, y)}{\partial y^2} \right)^2 dx dy.$$

- ▶ $\mathcal{L}^*(x, y)$, a function of point density in the new space.
- ▶ Adjust penalty:

$$\int \int_{\Omega} \mathcal{L}^*(x, y) \left(\left(\frac{\partial^2 f(x, y)}{\partial x^2} \right)^2 + \left(\frac{\partial^2 f(x, y)}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 f(x, y)}{\partial y^2} \right)^2 \right) dx dy.$$

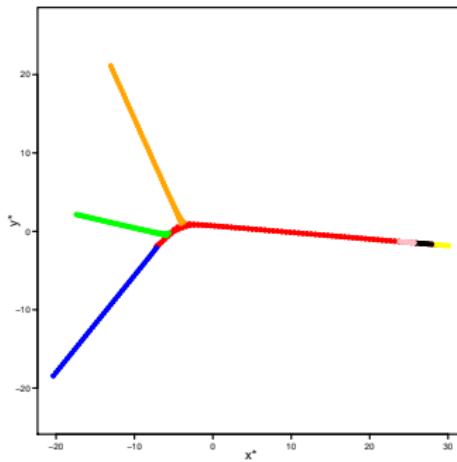
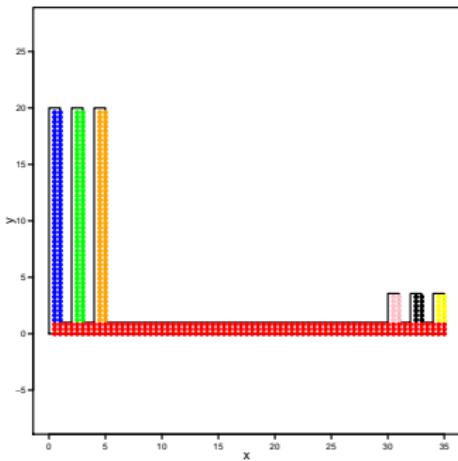
- ▶ Some improvements in MSE terms, doesn't get rid of all artefacts.

Calculating $\mathcal{L}^*(x, y)$



Ordering

- ▶ So, where are the artefacts coming from?
- ▶ Taking projections of n -dimensional space.
- ▶ Points can lose ordering.
- ▶ 3-D projection does better, but can we keep going?
(Nullspace gets BIG!)



Outline

Smoothing over complex regions

The problem

Smoothing with splines

Solutions

Multidimensional Scaling

Details

Finding the within-area distances

Simulation Results

Conclusions

Final comments

- ▶ Clearly, not that useful in practise.
- ▶ Better idea of what is required to make domain transformation work in general.

Criteria for domain transformation techniques:

- ▶ Minimize squashing. No crowding.
- ▶ Smooth mapping.
- ▶ Maintain order.
- ▶ Fast mapping.

The key: transform “just enough” but no more.

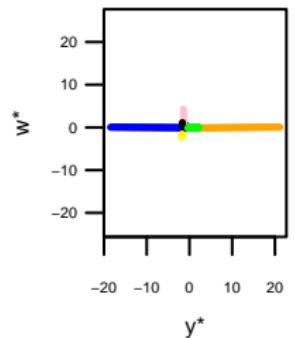
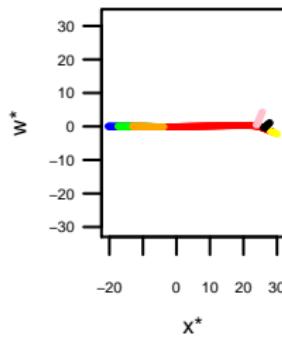
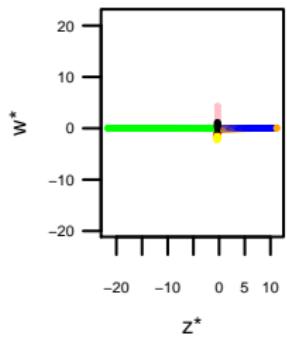
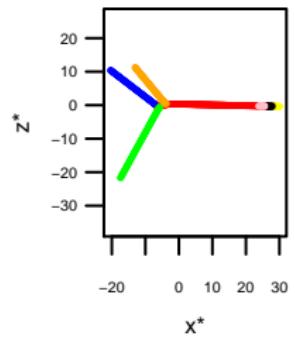
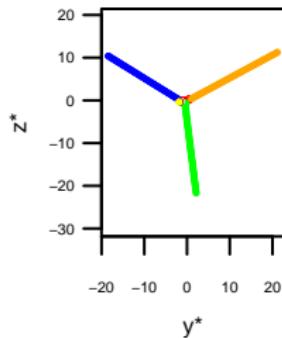
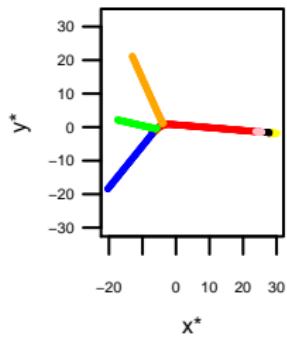
- ▶ Smoothing on manifolds?

References

- ▶ S.N. Wood, M.V. Bravington, and S.L. Hedley. *Soap film smoothing*. *JRSSB*, 2008
- ▶ H. Wang and M.G. Ranalli. *Low-rank smoothing splines on complicated domains*. *Biometrics*, 2007
- ▶ T.A. Driscoll and L.N. Trefethen. *Schwarz-Christoffel Mapping*. Cambridge, 2002
- ▶ T. Ramsay. *Spline smoothing over difficult regions*. *JRSSB*, 2002
- ▶ P.H.C. Eilers. *P-spline smoothing on difficult domains*. *University of Munich seminar*, 2006
- ▶ J.C. Gower. *Adding a point to vector diagrams in multivariate analysis*. *Biometrika*, 1968.

Slides available at <http://people.bath.ac.uk/dlm27>

Appendix - Ordering problems in 4-D



Appendix - A “new” algorithm for finding within-area distances

- ▶ Algorithm “bounces” around inside the polygon.
- ▶ Initial path is just the path around the edge, then iterate over two steps: *delete* and *alter*.
- ▶ Delete (iterating over all nodes):
 - ▶ If we can shorten the path by simply deleting a node, do that.
- ▶ Alter (iterating over all nodes):
 - ▶ If we can find a shorter sub-path by bouncing off the other “side” of the polygon, do that.