Machine Learning for Anatomical Neuroimaging

John Ashburner

Wellcome Trust Centre for Neuroimaging,
UCL Institute of Neurology,
12 Queen Square,
London WC1N 3BG,
UK.

"*The only relevant test of the validity of a hypothesis is comparison of prediction with experience.*"

Milton Friedman

Evidence-based Science

...also just known as "science".

- Researchers claim to find differences between groups. Do those findings actually discriminate?
- How can we most accurately diagnose a disorder from image data?
- Pharma wants biomarkers. How do we most effectively identify them?
- There are lots of potential imaging biomarkers. Which are most (cost) effective?

Pattern recognition provides a framework to compare data (or preprocessing strategy) to determine the most accurate approach.

Inter-subject Variability

Why focus on anatomy?

- Many medical applications involve understanding differences among individuals/populations.
- In image data, most of the differences we can see are anatomical in nature.
- Understanding growth and development requires us to look at growth and development (anatomy).

Bayesian approaches may be better for clinical applications

- Deals with different priors.
  - Consider a method with 90% sensitivity and specificity.
  - Consider using this to screen for a disease afflicting 1% of the population.
  - On average, out of 100 people there would be 10 wrongly assigned to the disease group.
  - A positive diagnosis suggests only about a 10% chance of having the disease.

$$P(\text{Disease}|\text{Pred}+) = \frac{P(\text{Pred}+|\text{Disease})P(\text{Disease})}{P(\text{Pred}+|\text{Disease})P(\text{Disease})+P(\text{Pred}+|\text{Healthy})P(\text{Healthy})}$$

- Better decision-making by accounting for utility functions.
- Confidence may differ from subject to subject.

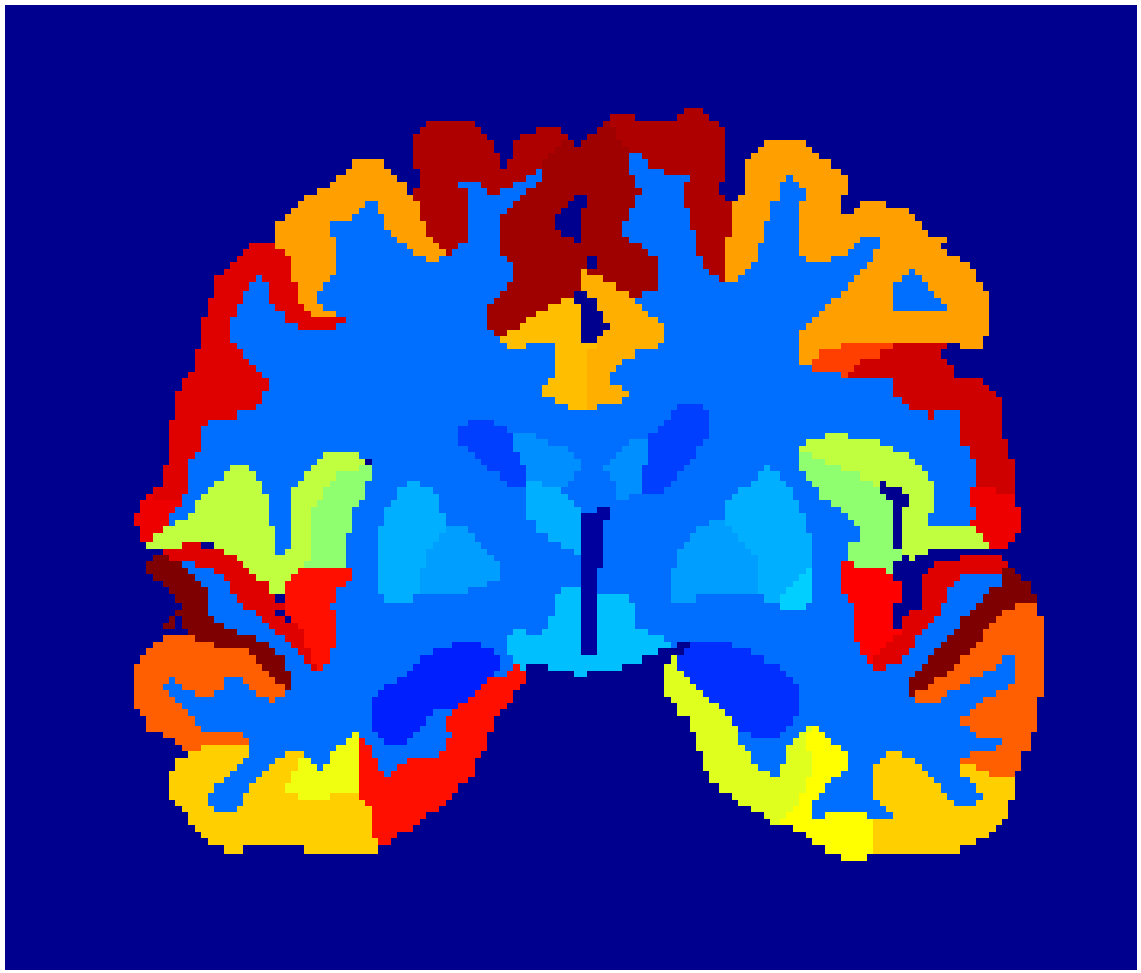
Feature engineering

First-timers are often surprised by how **little time in a machine learning project is spent actually doing machine learning**. *But it makes sense if you consider how time-consuming it is to gather data, integrate it, clean it and pre-process it, and how much* **trial and error can go into feature design**. *Also, machine learning is not a one-shot process of building a data set and running a learner, but rather an iterative process of running the learner, analyzing the results, modifying the data and/or the learner, and repeating. Learning is often the quickest part of this, but that's because we've already mastered it pretty well!* **Feature engineering is more difficult because it's domain-specific, while learners can be largely general-purpose.** *However, there is no sharp frontier between the two, and this is another reason* **the most useful learners are those that facilitate incorporating knowledge**.

Domingos, Pedro. "A few useful things to know about machine learning." Communications of the ACM 55, no. 10 (2012): 78-87.

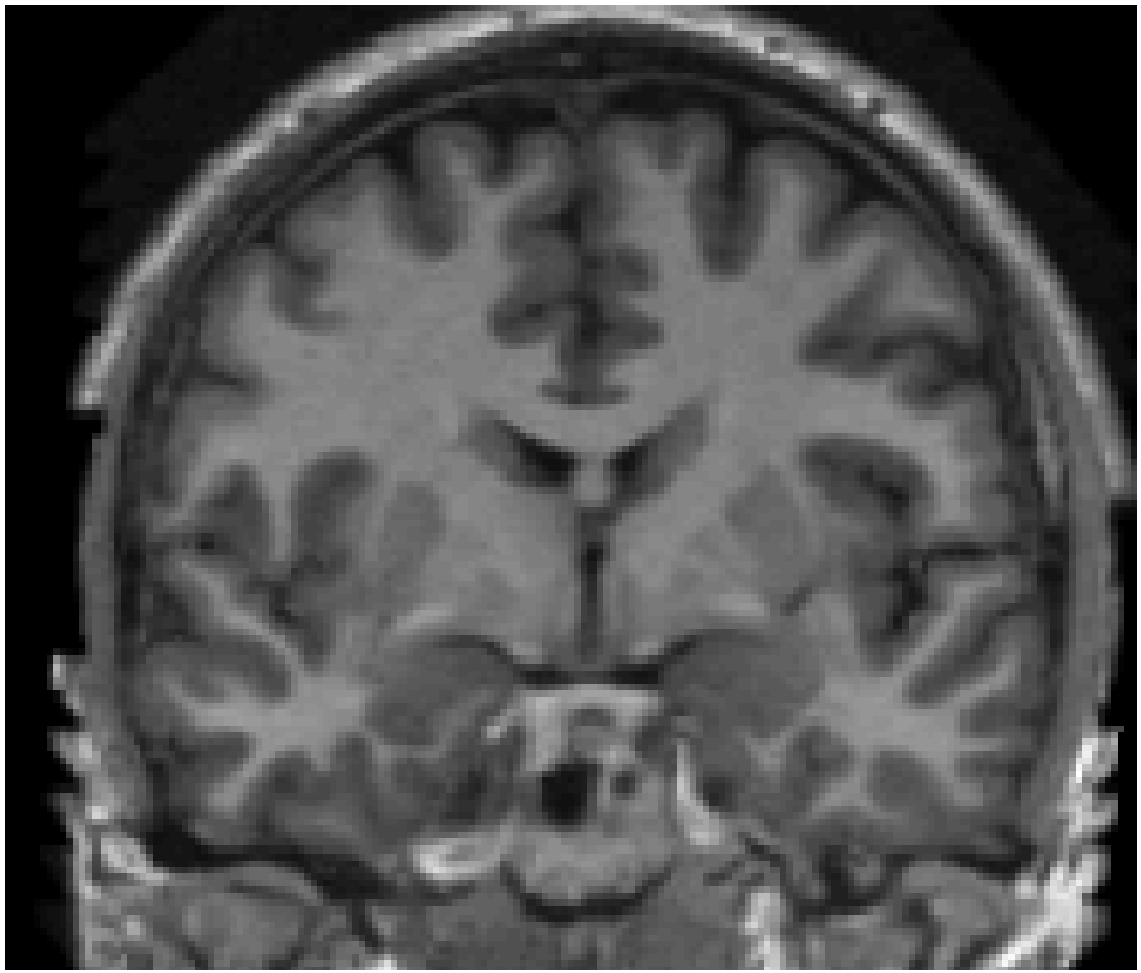
Region volumes

Label propagation or other methods can be used to subdivide brain into regions.



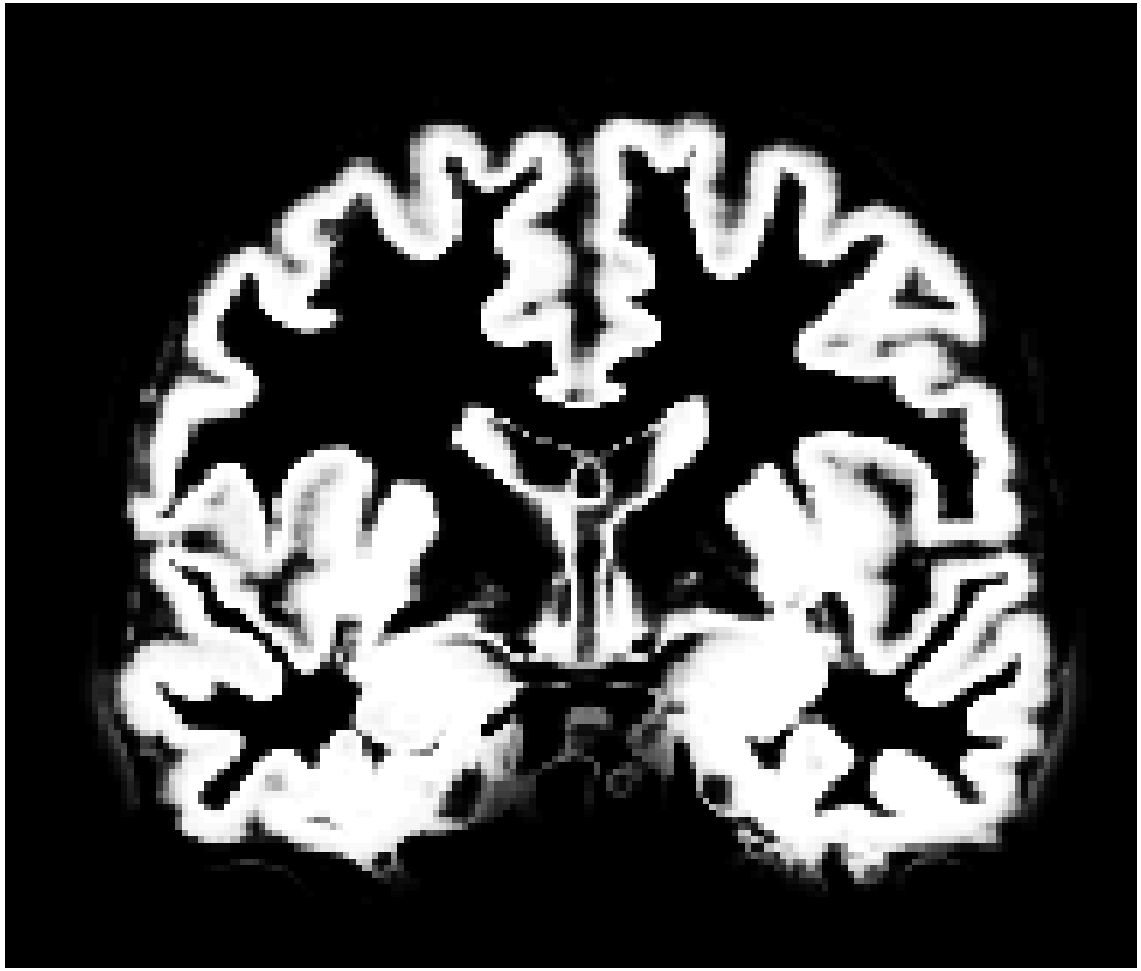
Pixel values

Raw pixel data could be another option.
Data needs to be "spatially normalised" (and possibly skull-stripped).
Results may not generalise well to data from other scanners.



Tissue maps

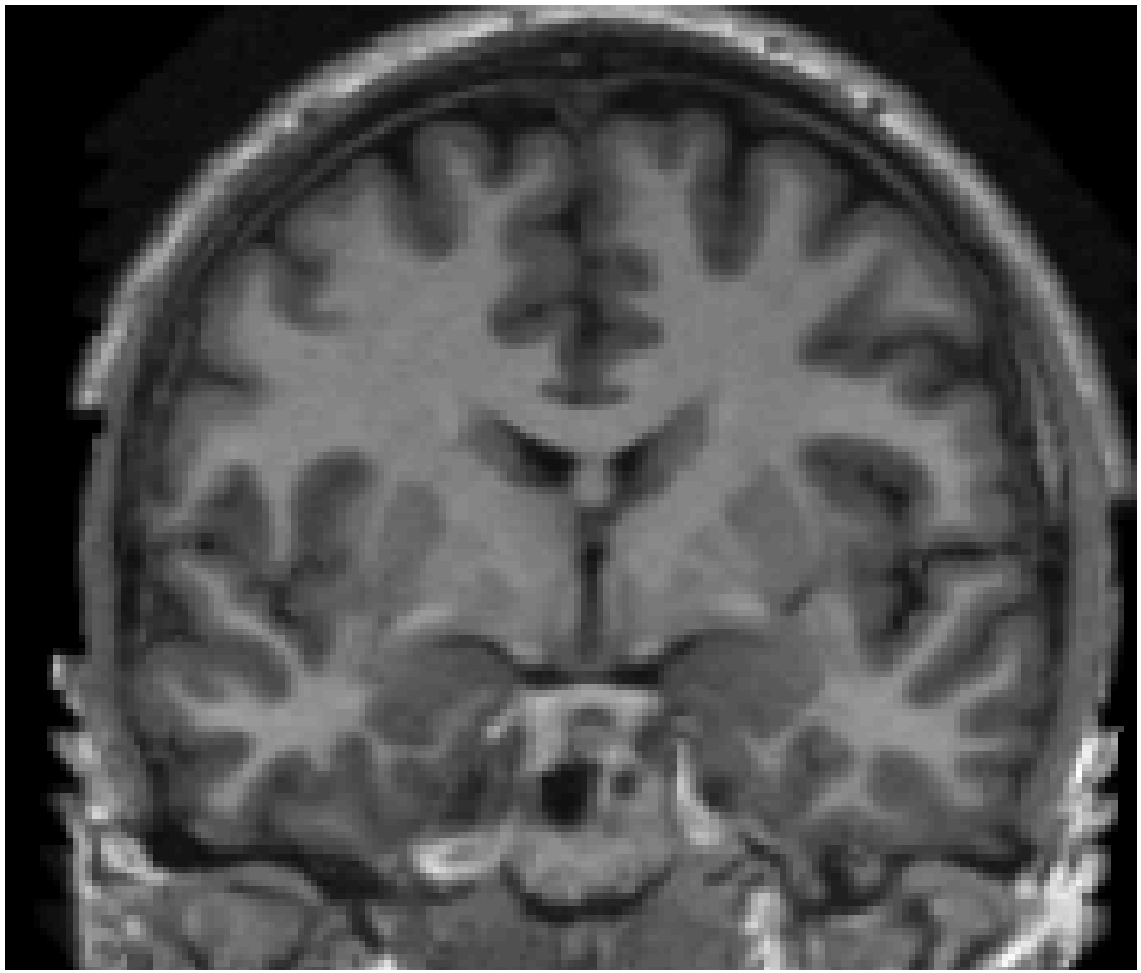
Grey matter maps can work fairly well.
Data needs to be "spatially normalised".
Many neurological problems show up as grey matter atrophy.



Other features

Other features include:

- Cortical thickness.
- Shape features.
- PCA/ICA weights.
- Lesion maps.
- etc



No Free Ducklings

No Free Lunch theorem says that learning is impossible without prior knowledge.
`http://en.wikipedia.org/wiki/No_free_lunch_in_search_and_optimization`

Ugly Duckling theorem says that things are all equivalently similar to each other without prior knowledge.
`http://en.wikipedia.org/wiki/Ugly_duckling_theorem`

By Ryan Ebert from Portland, US (Flickr) [CC BY 2.0], via Wikimedia Commons.
https://creativecommons.org/licenses/by/2.0/

What prior knowledge do we have about the variability among people that can be measured using MRI?
How do we use this knowledge?

Different ways of measuring distances

David Mumford

Mathematician

David Bryant Mumford is an American mathematician known for distinguished work in algebraic geometry, and then for research into vision and pattern theory. He won the Fields Medal and was a MacArthur Fellow. Wikipedia

Born: June 11, 1937 (age 76), Worth village, West Sussex, Crawley

Children: Steve Mumford

Education: Phillips Exeter Academy, Harvard University

Awards: Fields Medal, Wolf Prize in Mathematics, MacArthur Fellowship, The Shaw Prize in Mathematical Sciences, National Medal of Science for Mathematics and Computer Science

Empirical Statistics and Stochastic Models for Visual Signals

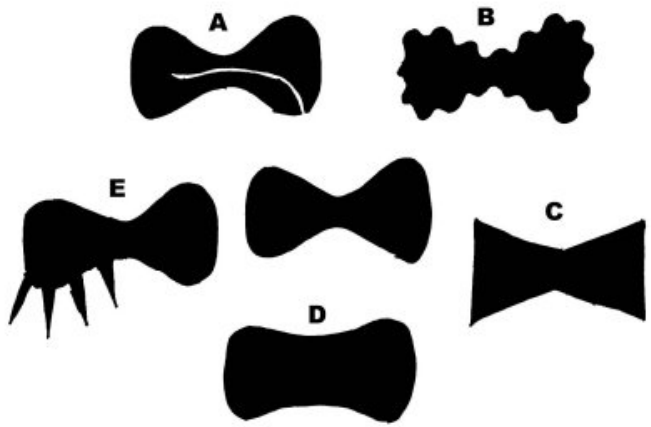


Figure 1.11 Each of the shapes A,B,C,D and E is similar to the central shape, but *in different ways*. Different metrics on the space of shape bring out these distinctions.

Kernel Matrices

Linear kernel matrices are often computed from the raw features:

$$\mathbf{K} = \mathbf{X}\mathbf{X}^T$$

A simple spatial feature selection may be considered as the following, where $\mathbf{\Sigma}_0$ is a (scaled) diagonal matrix of ones and zeros:

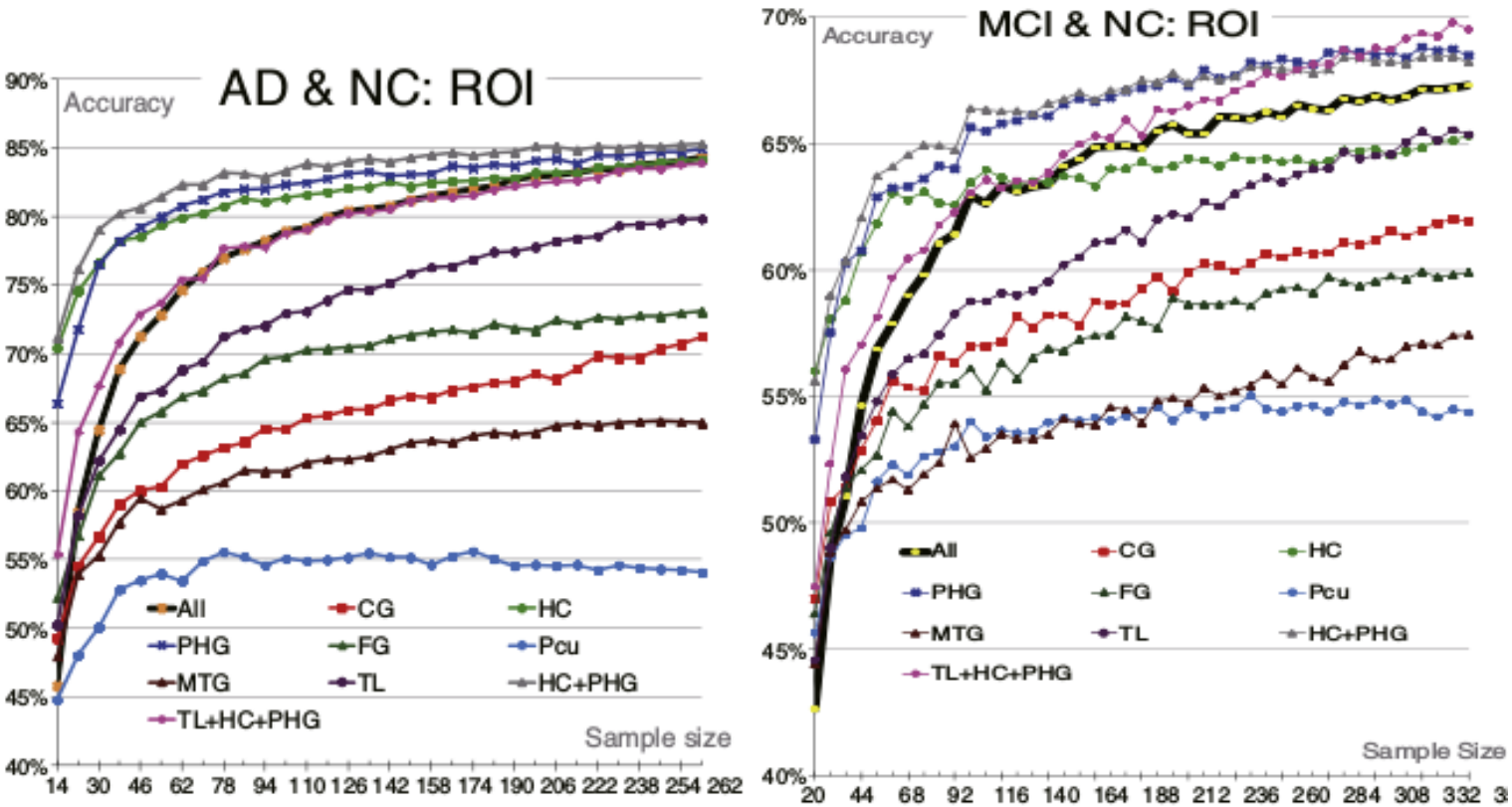
$$\mathbf{K} = \mathbf{X}\mathbf{\Sigma}_0\mathbf{X}^T$$

$\mathbf{\Sigma}_0$ may be more complicated, for example encoding spatial smoothing, high-pass filtering or any number of other things.

Prior knowledge about brain regions involved

- The best way would be to augment the training data with data from previous studies.
- Lack of data-sharing means this is generally not possible, so we need to extract information from publications.
- The neuroimaging literature is mostly blobs.
- These give pointers about how best to weight the data ($\mathbf{\Sigma}_0 = diag(\mathbf{s}), s_i \in \mathbb{R}^+$).

Weighting suspected regions more heavily



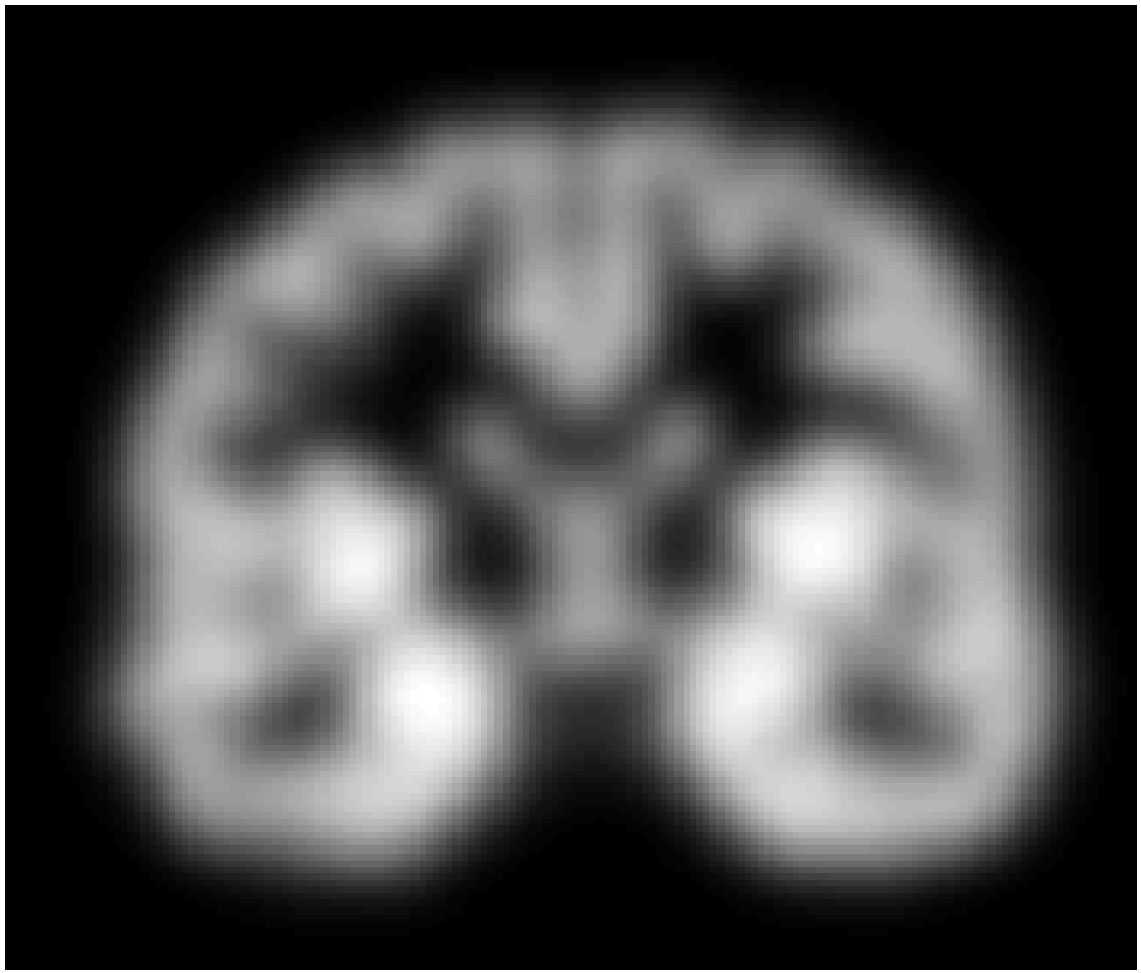
Chu et al. "Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images". NeuroImage 60:59–70 (2012).

Smoothing can help

If we know that higher frequency signal is more likely to be noise.

$$\mathbf{K} = \mathbf{X}\boldsymbol{\Sigma}_0\mathbf{X}^T$$

$\boldsymbol{\Sigma}_0$ no longer diagional.



Dimensionality $\neq$ number of voxels

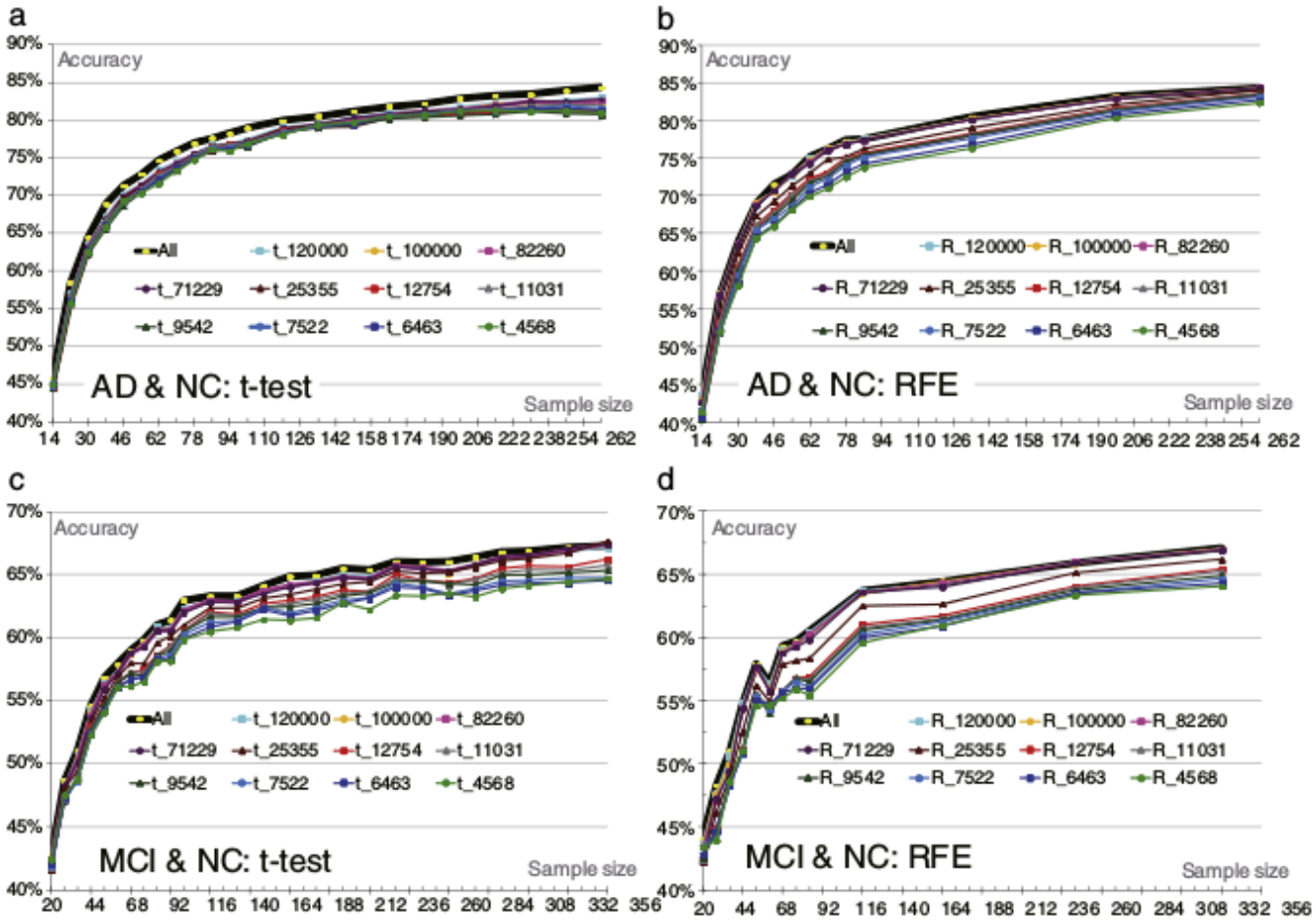
- Lots of effort on data-driven feature selection methods.
  - Involves estimating $\mathbf{\Sigma}_0 = diag(\mathbf{s}), s_i \in \{0, w\}$, where $w \in \mathcal{R}^+$.
  - Lots of parameters needed to achieve this.
- Many papers claim excellent results.
- Little evidence to suggest that most voxel-based feature selection methods help.
  - Little or no increase in predictive accuracy.
  - Commonly perceived as being more "interpretable".

"Data-driven feature selection"

"In our evaluation, two methods included a feature selection step: Voxel-STAND and Voxel-COMPARE. Overall, these methods did not perform substantially better than simpler ones... ... A more robust way to decrease the dimensionality of the features way would be to use more prior knowledge of the disease."

Cuingnet et al. "Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database". NeuroImage 56(2):766–781 (2011).

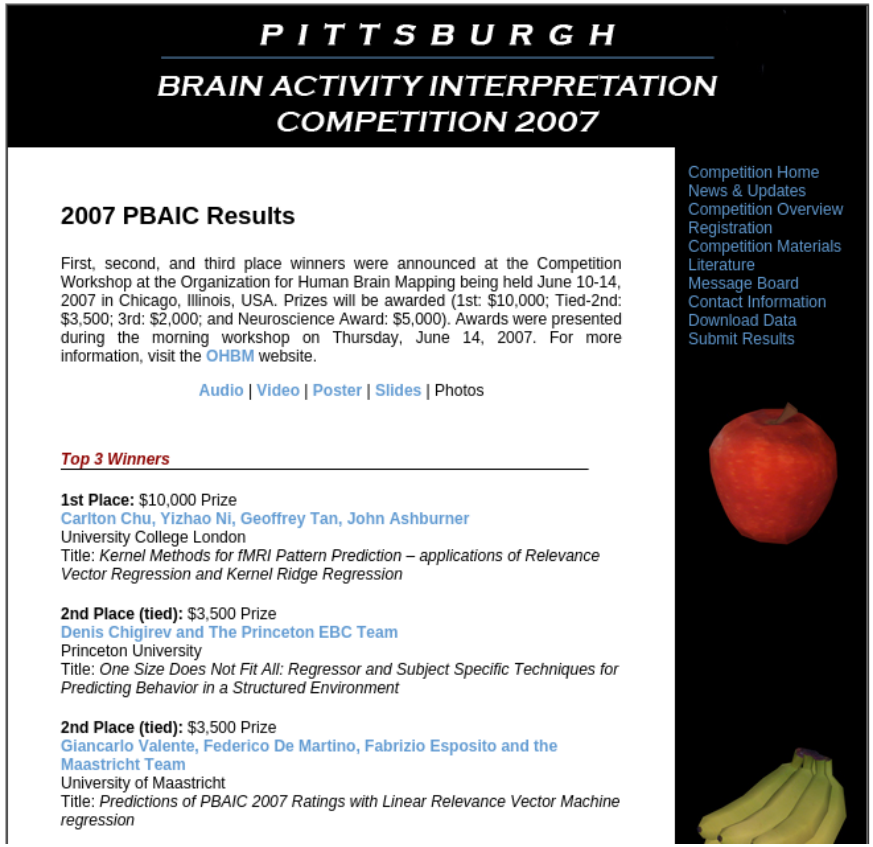
"Data-driven feature selection"



Chu et al. "Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images". NeuroImage 60:59–70 (2012).

"Data-driven feature selection"

...did not help the winning entry.



PITTSBURGH

BRAIN ACTIVITY INTERPRETATION COMPETITION 2007

Competition Home
News & Updates
Competition Overview
Registration
Competition Materials
Literature
Message Board
Contact Information
Download Data
Submit Results

2007 PBAIC Results

First, second, and third place winners were announced at the Competition Workshop at the Organization for Human Brain Mapping being held June 10-14, 2007 in Chicago, Illinois, USA. Prizes will be awarded (1st: $10,000; Tied-2nd: $3,500; 3rd: $2,000; and Neuroscience Award: $5,000). Awards were presented during the morning workshop on Thursday, June 14, 2007. For more information, visit the OHBM website.

Audio | Video | Poster | Slides | Photos

Top 3 Winners

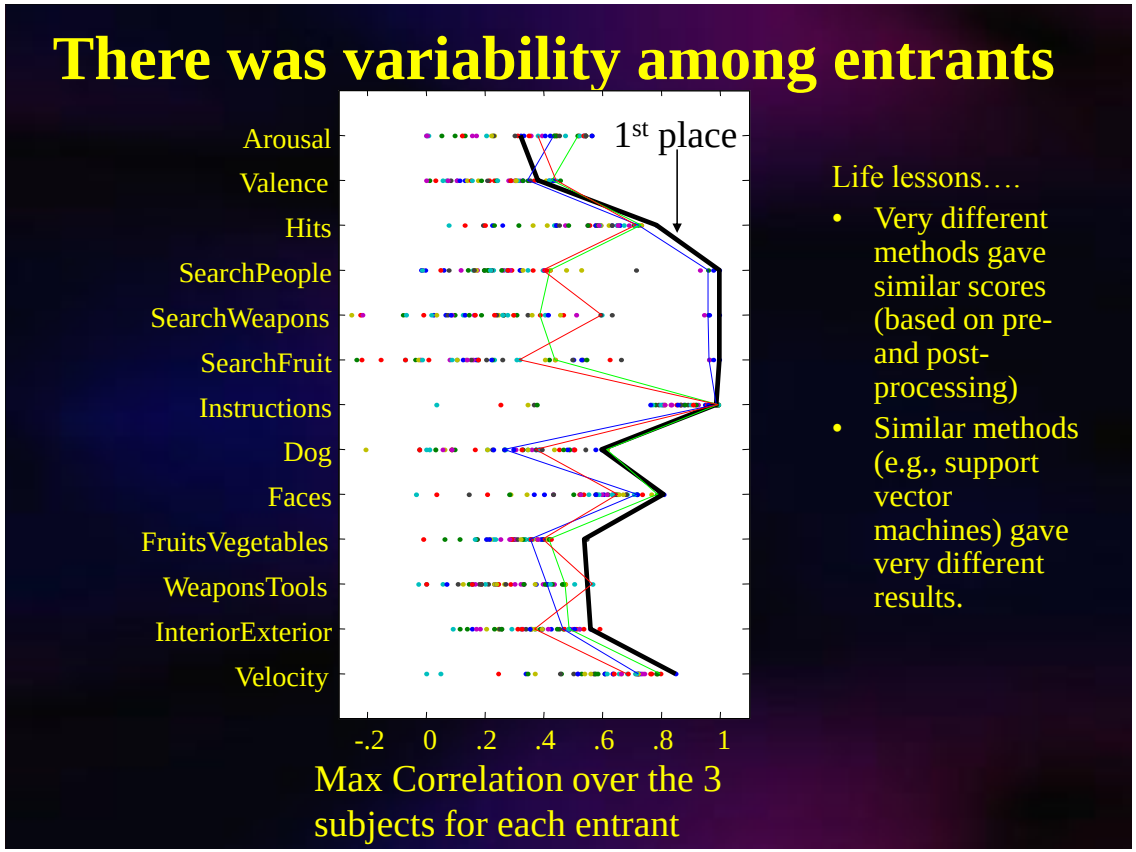
1st Place: $10,000 Prize
Carlton Chu, Yizhao Ni, Geoffrey Tan, John Ashburner
University College London
Title: *Kernel Methods for fMRI Pattern Prediction – applications of Relevance Vector Regression and Kernel Ridge Regression*

2nd Place (tied): $3,500 Prize
Denis Chigirev and The Princeton EBC Team
Princeton University
Title: *One Size Does Not Fit All: Regressor and Subject Specific Techniques for Predicting Behavior in a Structured Environment*

2nd Place (tied): $3,500 Prize
Giancarlo Valente, Federico De Martino, Fabrizio Esposito and the Maastricht Team
University of Maastricht
Title: *Predictions of PBAIC 2007 Ratings with Linear Relevance Vector Machine regression*

http://www.lrdc.pitt.edu/ebc/2007/2007.html

"Data-driven feature selection"



http://www.lrdc.pitt.edu/ebc/2007/2007.html

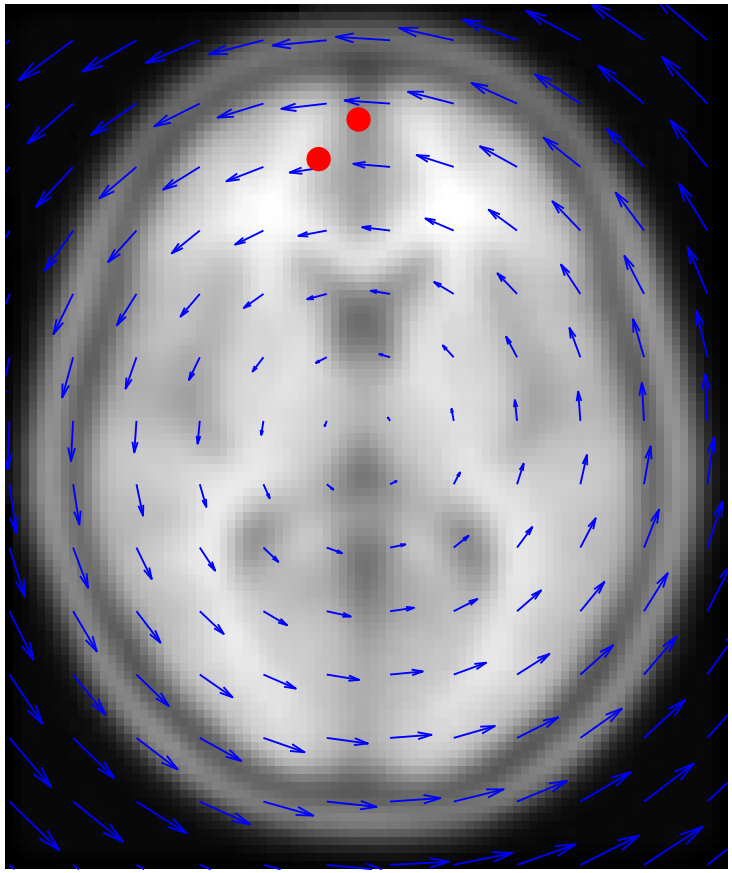
Modelling the nonlinearities

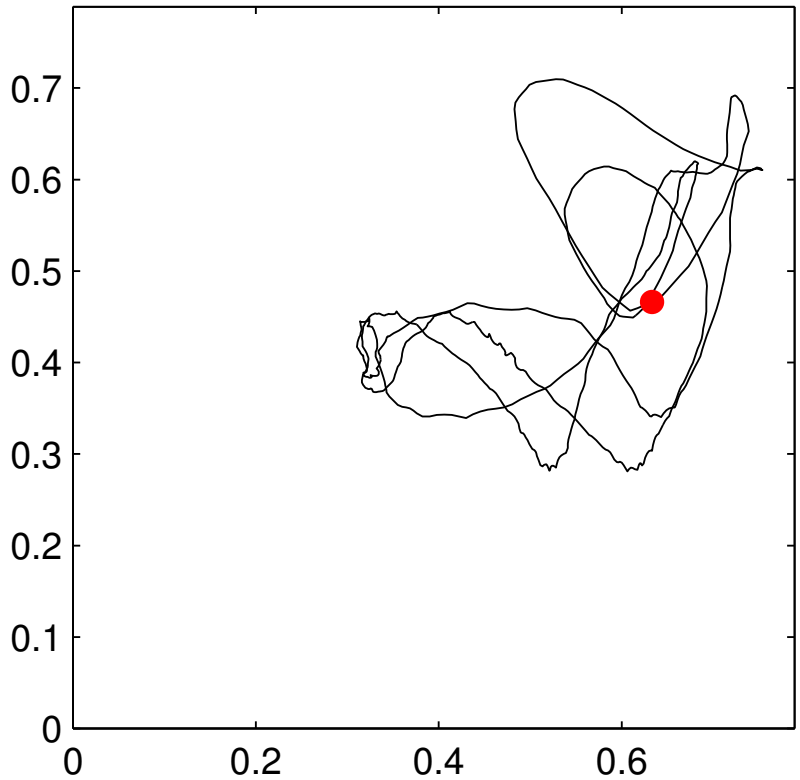
Instead of using nonlinear pattern recognition methods...

- Capture nonlinearities by appropriate preprocessing.
- Allows nonlinear effects to be modelled by a linear classifier.
- Gives more interpretable characterisations of differences.
- May lead to more accurate predictions.

Transformed images fall on manifolds

Rotating an image leads to points on a manifold.

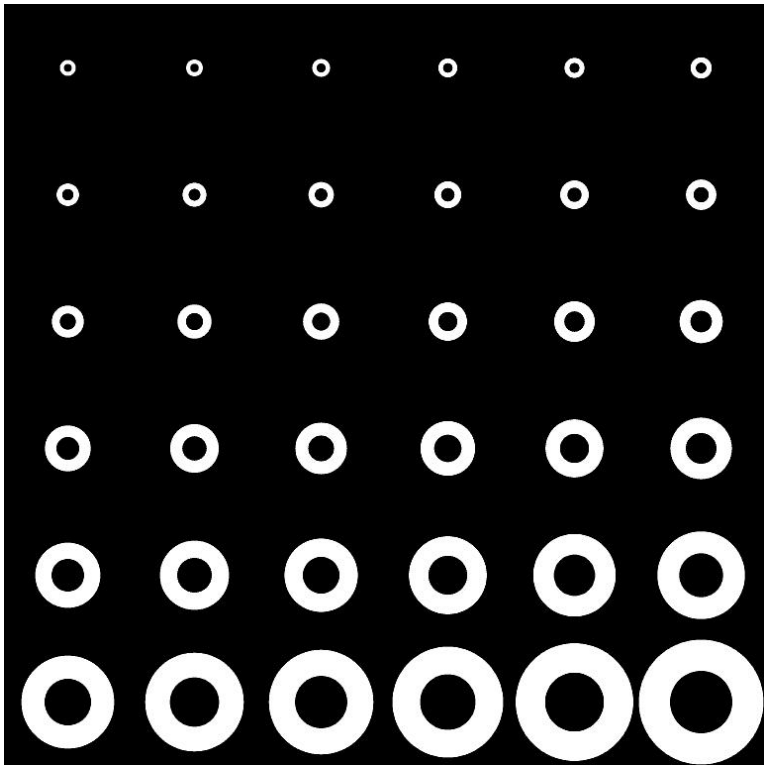




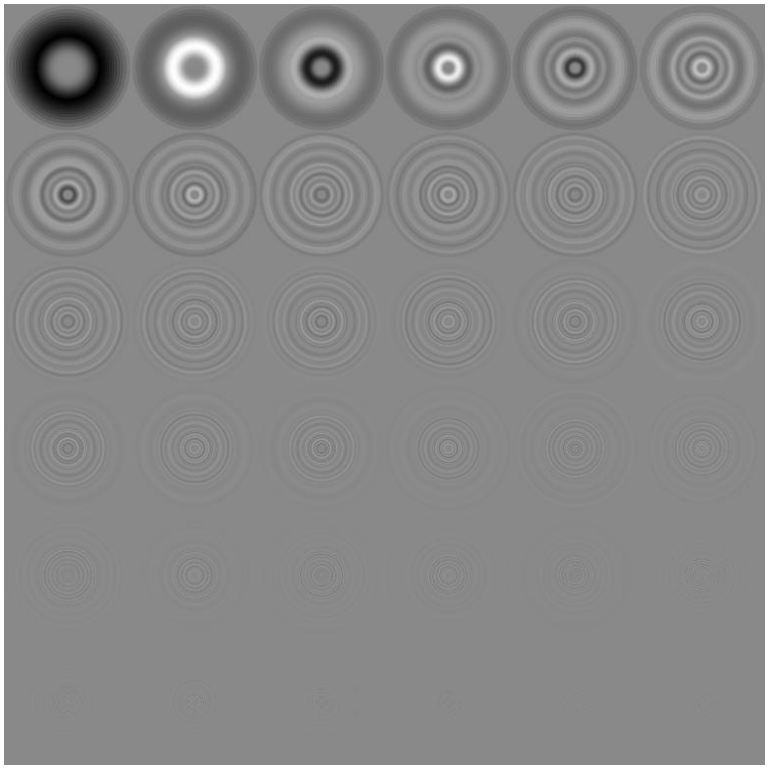
Rigid-body motion leads to a 6-dimensional manifold (not shown).

One mode of geometric variability

Simulated images



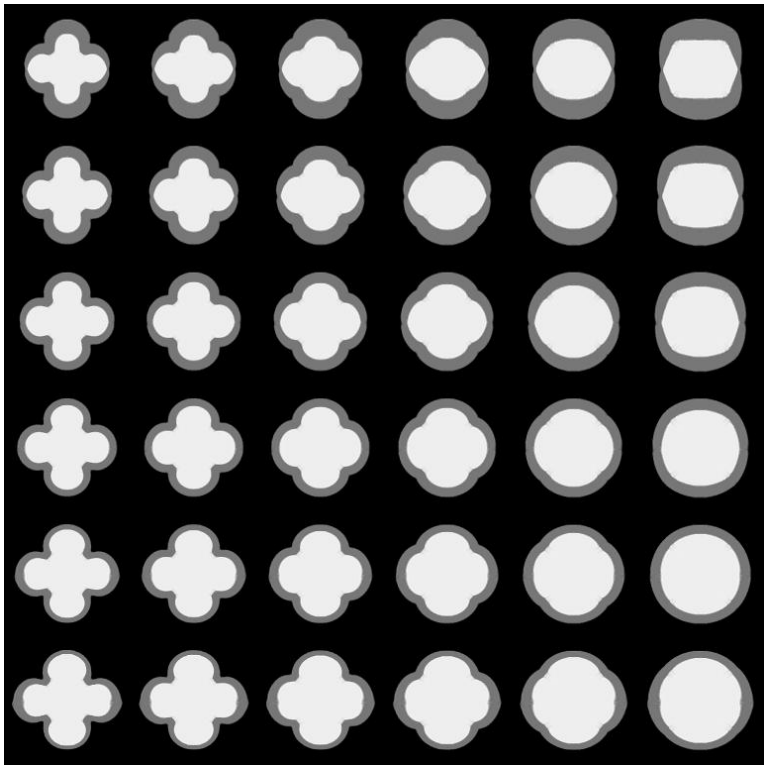
Principal components



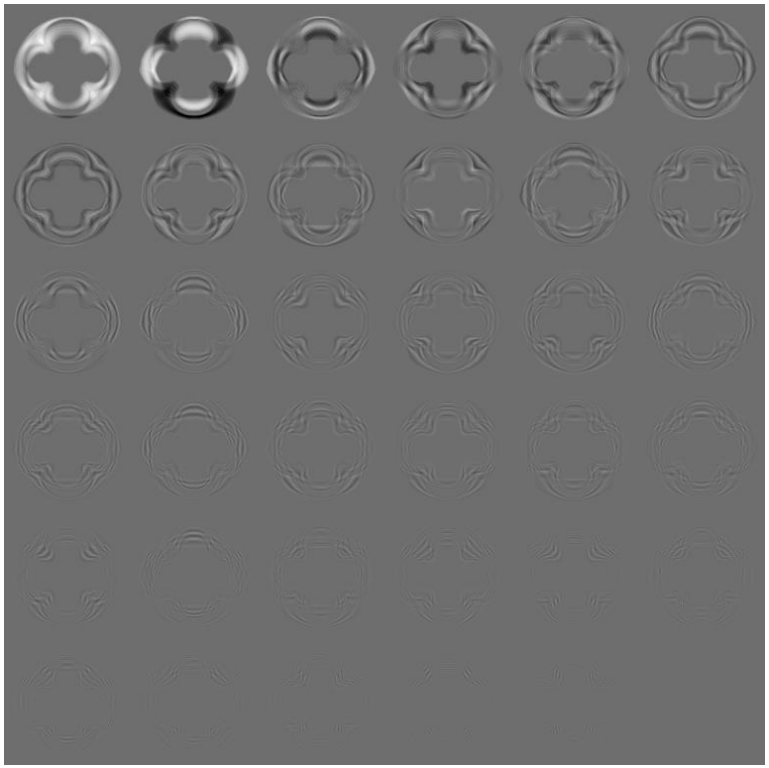
A suitable model would reduce this variability to a single dimension.

Two modes of geometric variability

Simulated images



Principal components



A suitable model would reduce this variability to two dimensions.

"Traditional" morphometrics

Traditional morphometrics analyzes measurements of size (lengths, widths, masses, angles, ratios and areas).
Early analysis methods were often univariate, performing statistical analysis on each feature in isolation.

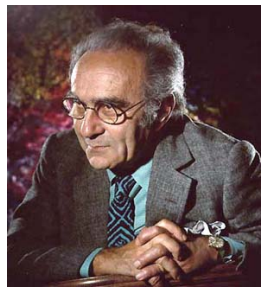
Wikipedia contributors, "Morphometrics," Wikipedia, The Free Encyclopedia, `http://en.wikipedia.org/w/index.php?title=Morphometrics&oldid=641987541` (accessed March 26, 2015).

Criticism of univariate approaches

"This unhappy result can be traced to the piecemeal tests which have hitherto been used. A bone or a tooth is a unit ; it is not a discrete assembly of independent measurements."

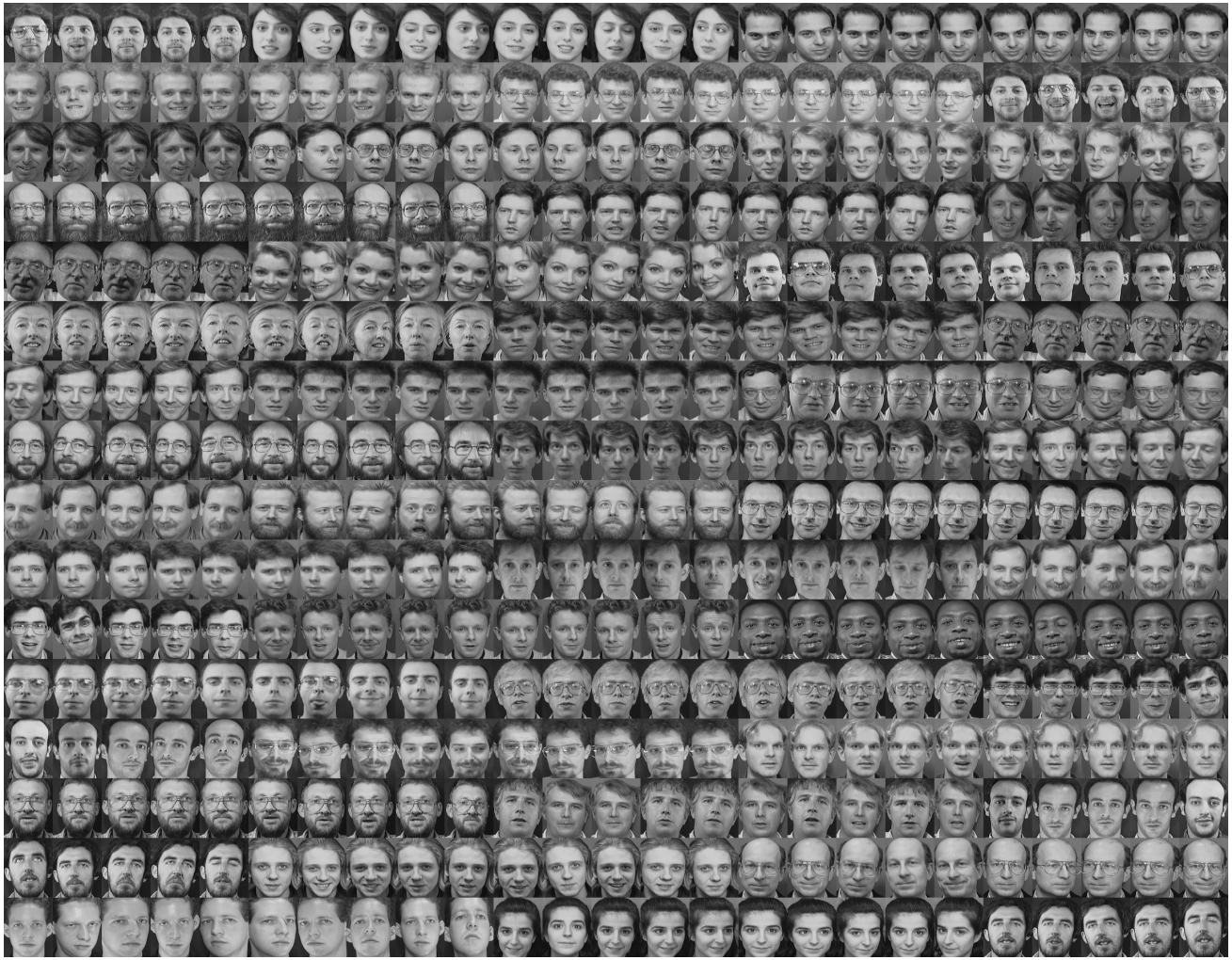
— Jacob Bronowski & W.M. Long (Nature, 1951)

"The right statistical method must treat the set of variates as a single coherent matrix ; and this is, in fact, the technique of multivariate analysis."



— Jacob Bronowski & W.M. Long (Nature, 1951)

Biological variability is multivariate



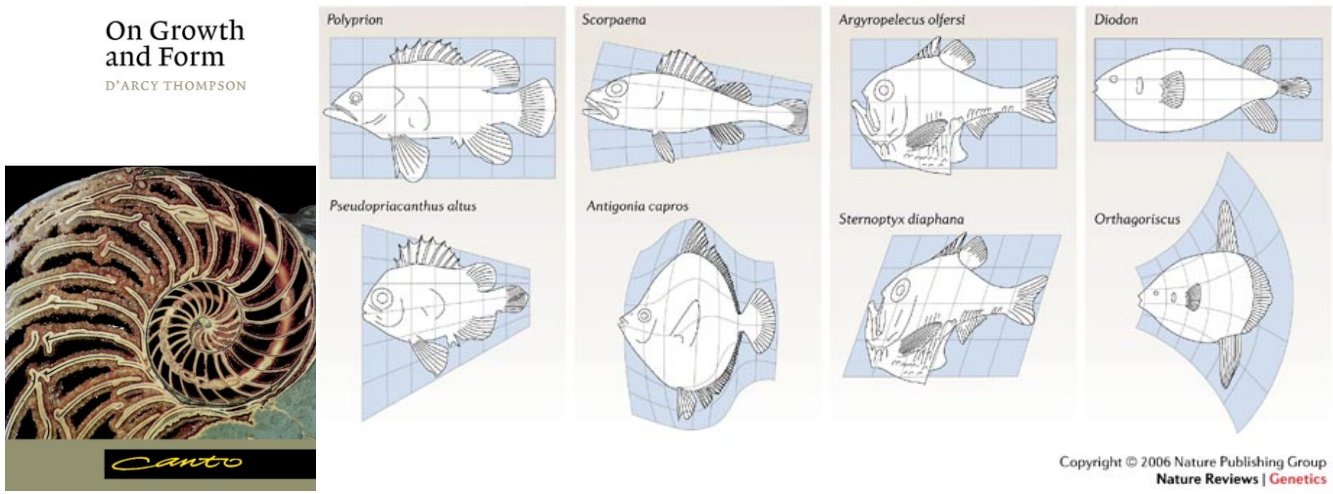
Geometric morphometrics

"We are now in the midst of a **revolution in morphometric methodology**. *The new approaches are more effective in capturing information about the shape of an organism and result in* **more powerful statistical procedures** *for testing for differences in shape. They are also* **more effective in enabling a researcher to visualize differences** *in shape and in suggesting simple traditional measurements that could be used in future studies."*

FJ Rohlf & LF Marcus. "A Revolution in Morphometrics". Trends in Ecology & Evolution 8.4: 129-132 (1993).

D'Arcy Thompson's approach

"...diverse and dissimilar fishes can be referred as a whole to identical functions of very different co-ordinate systems..."



Thompson, D'Arcy. "On Growth and Form." Cambridge University Press 1917.

Geometric morphometrics

1. Data are recorded to capture the geometry in the form of 2D or 3D **coordinates of landmark points**.
2. Geometric relationships among landmarks are not inherent in the raw coordinates themselves. **The relationship among points is captured by fitting an appropriate function to them** in 2 or 3D.
3. **The analyses are designed to indicate directions of maximum variation** and hence may suggest which conventional variables one should emphasize in verbal descriptions of the results.
4. Displays of the results of the analyses are emphasized, using **differences or changes that can be shown on pictorial representations** of the organisms studied.

FJ Rohlf & LF Marcus. "A Revolution in Morphometrics". Trends in Ecology & Evolution 8.4: 129-132 (1993).

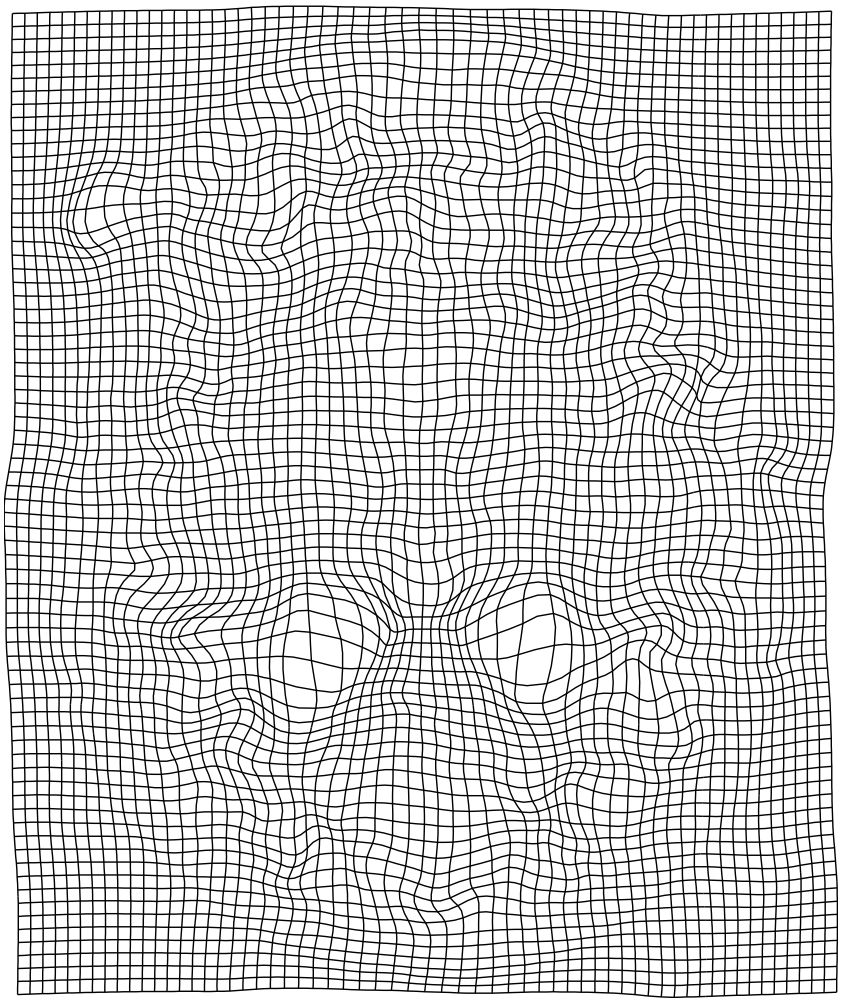
Geometric morphometrics

Classic multivariate statistical methods used:

- Procrustes analysis
- Principal Component Analysis (PCA)
- Canonical Correlation Analysis (CCA)
- Multivariate ANalysis of COVAriance (MANCOVA)
- Discriminant functions

These appraches could be augmented by pattern recognition and other machine learning techniques.

Encoding geometry



- Manual definition of landmarks is laborious, subjective and not very reproducible.
- Relatively few landmarks in the brain.
- Image registration is easier.
- Can apply the tools of geometric morphometrics (statistical shape analysis) to automatically estimated deformations.

Logarithmic relationships

PROBLEMS OF RELATIVE GROWTH

CHAPTER I

CONSTANT DIFFERENTIAL GROWTH-RATIOS

Huxley. "Problems of relative growth." Methuen & co, London (1932).

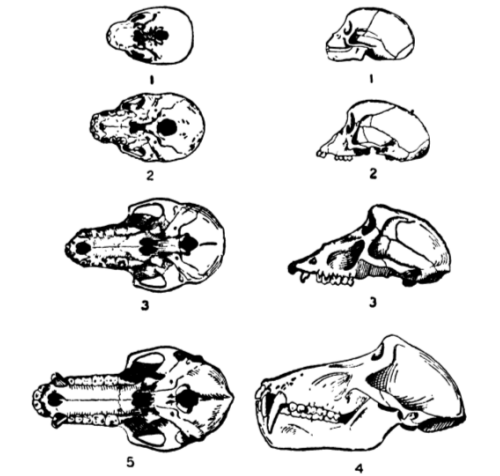


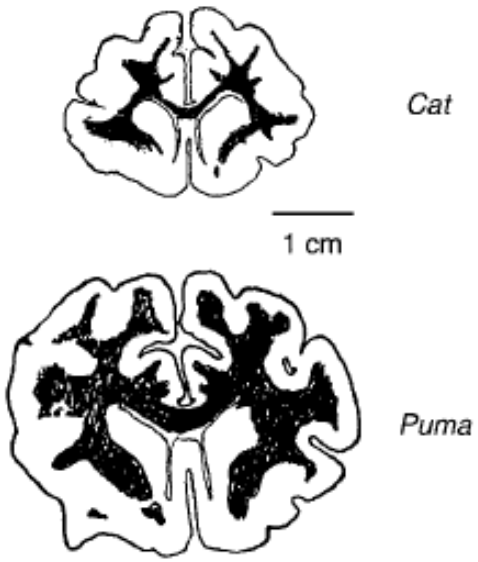
Fig. 10.—Baboon skulls of various sizes, to show the increase in relative size of facial region with absolute size of skull.
1, new-born; 2, juvenile (with milk dentition); 3, adult female; 4, adult male.

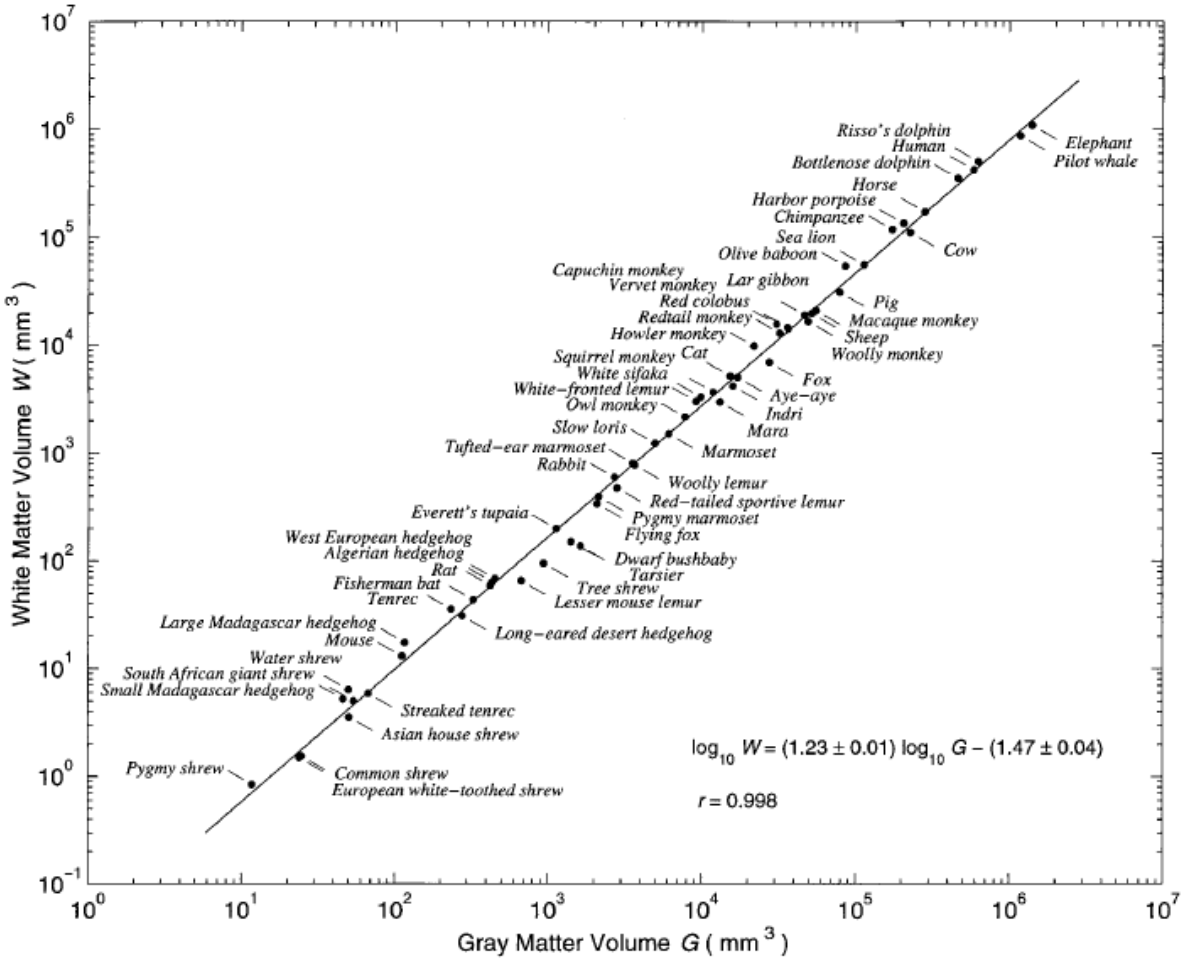
Julian Huxley demonstrated logarithmic relationships between magnitude variables (eg height, weight, length, area, volume).

$$\log y = a \log x + \log k$$

"the logarithmic method of plotting brings into true relief an important point which is entirely obscured by the usual method of plotting on the absolute scale – namely the fact that growth is concerned essentially with the multiplication of living substance."

Logarithmic relationships

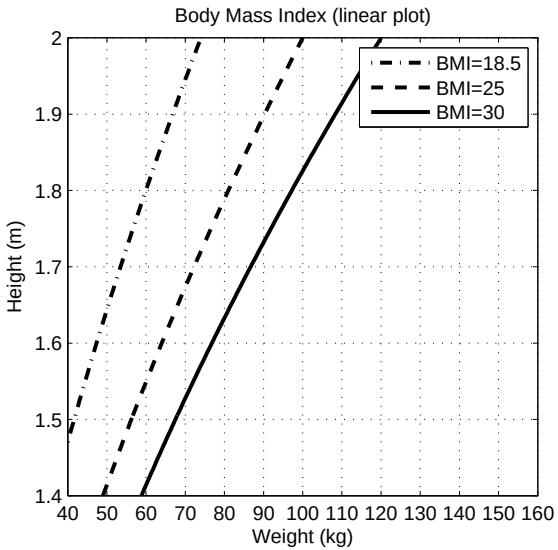


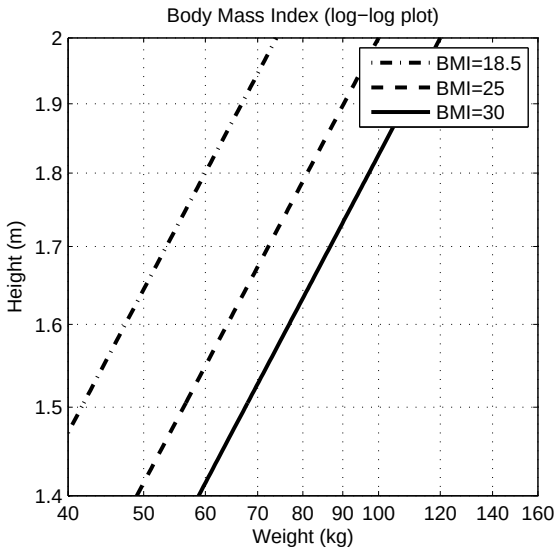


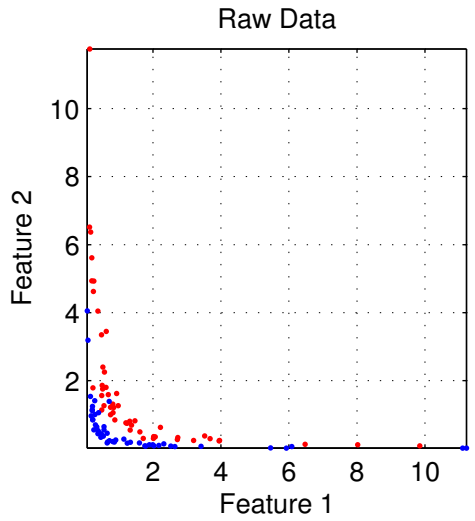
Zhang and Sejnowski. "A universal scaling law between gray matter and white matter of cerebral cortex." Proceedings of the National Academy of Sciences 97(10):5621–5626 (2000).

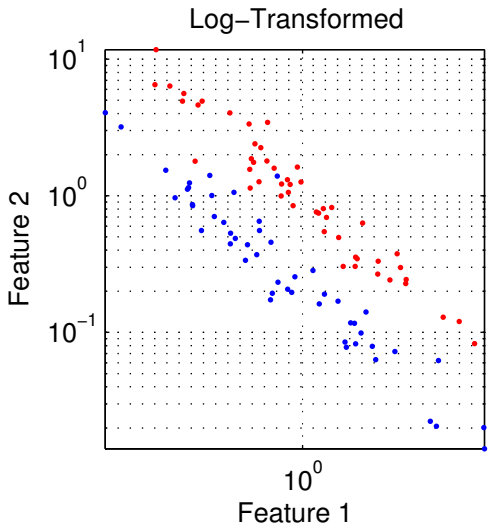
Logarithmic relationships

Preprocess to obtain features that behave more linearly.









Exponentials

There are many types of exponentials and their inverses.

Non-Euclidean geometry

- Distances are not always measured along a straight line.
- Sometimes we want distances measured on a manifold.
- Shortest path on a manifold is along a *geodesic*.

Linear trajectory

Nonlinear trajectory

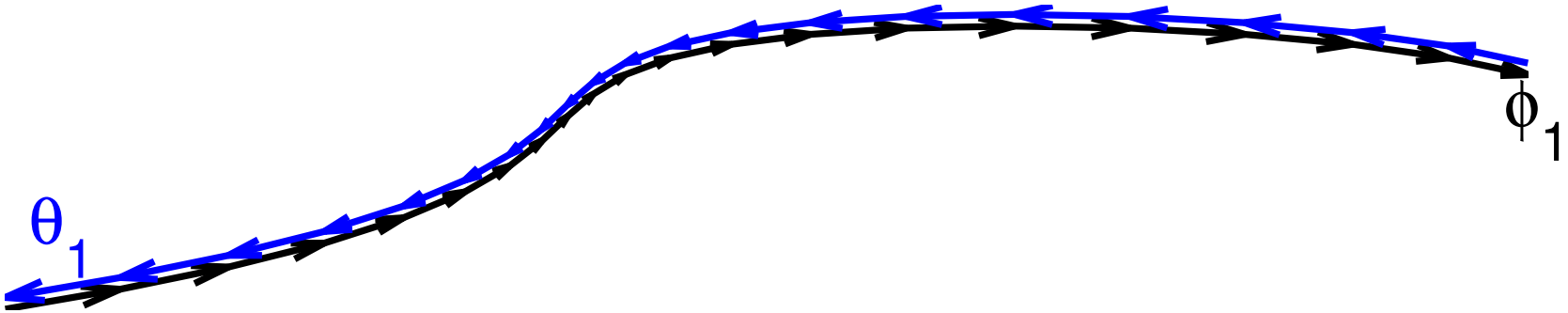
Metric distances

Distances should satisfy the properties of a *metric*:

1. $d(\mathbf{x}, \mathbf{y}) \geq 0$ (non-negativity)
2. $d(\mathbf{x}, \mathbf{y}) = 0$ if and only if $\mathbf{x} = \mathbf{y}$ (identity of indiscernibles)
3. $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$ (symmetry)
4. $d(\mathbf{x}, \mathbf{z}) \leq d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z})$ (triangle inequality).

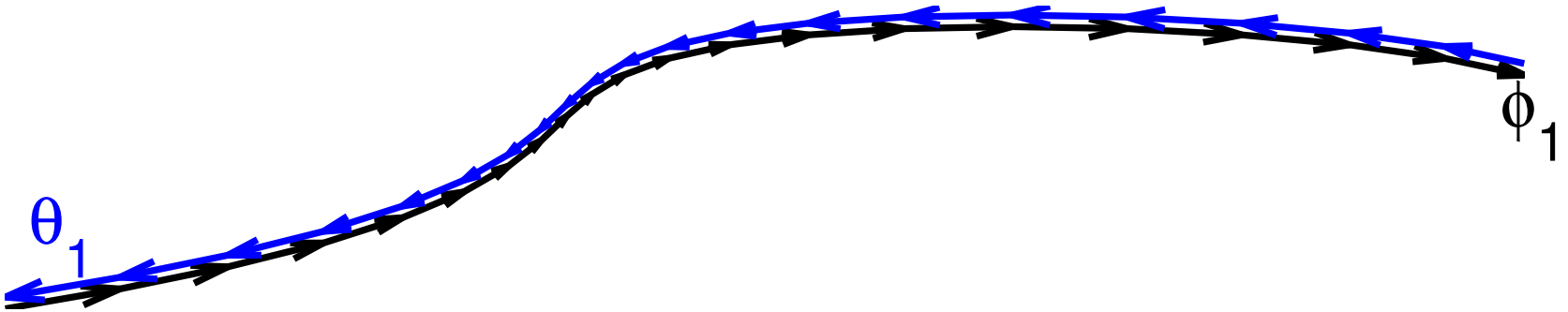
Satisfying (3) requires inverse-consistent image registration.
Satisfying (4) requires a specific class of image registration models.

Computing a metric distance



Decompose a curved path into a series of short line segments, and add the lengths of the segments together.

Computing large deformations



We can consider a large deformation as the composition of a series of small deformations:

$$\varphi_1 = \left(\mathrm{id} + \frac{\mathbf{v}_{t_{N-1}}}{N}\right) \circ \left(\mathrm{id} + \frac{\mathbf{v}_{t_{N-2}}}{N}\right) \circ \ldots \circ \left(\mathrm{id} + \frac{\mathbf{v}_{t_1}}{N}\right) \circ \left(\mathrm{id} + \frac{\mathbf{v}_0}{N}\right)$$

The inverse of the deformation can be computed from:

$$\vartheta_1 = \left(\mathrm{id} - \frac{\mathbf{v}_0}{N}\right) \circ \left(\mathrm{id} - \frac{\mathbf{v}_{t_1}}{N}\right) \circ \ldots \circ \left(\mathrm{id} - \frac{\mathbf{v}_{t_{N-2}}}{N}\right) \circ \left(\mathrm{id} - \frac{\mathbf{v}_{t_{N-1}}}{N}\right)$$

Metric distances from large deformations

By modelling trajectories as piecewise linear, distances can be computed by adding the distances from the small deformations:

$$d = \frac{1}{N} \sum_{n=0}^{N-1} ||\mathbf{L}\mathbf{v}_{t_n}||$$

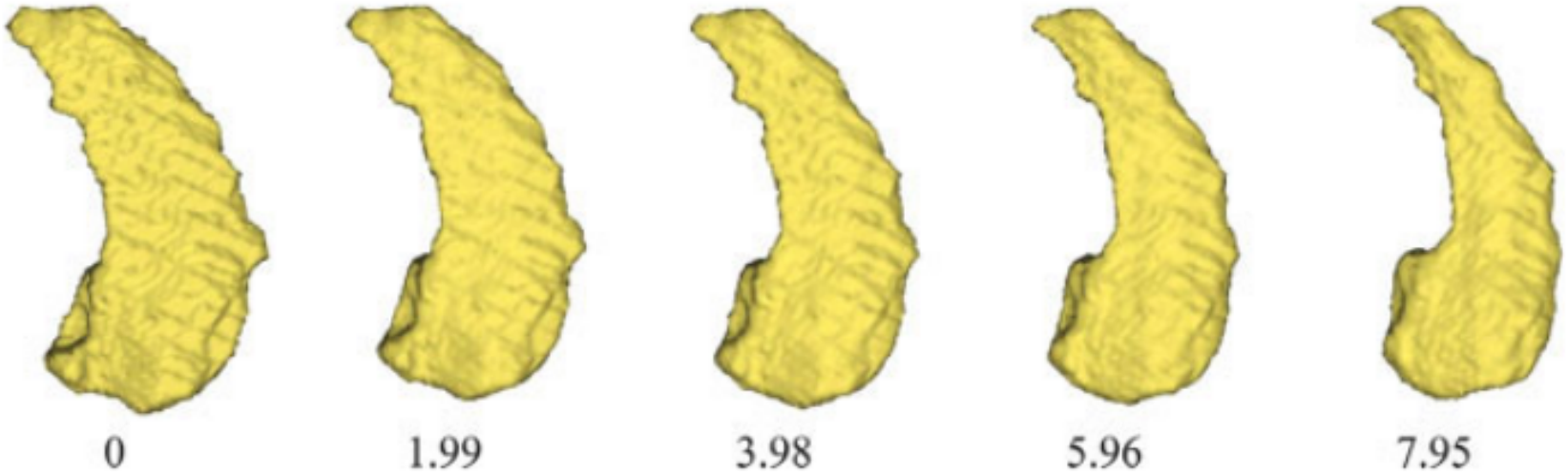
If N approaches infinity (and we use small deformations of $\mathrm{id} + \frac{1}{N}\mathbf{v}_t$), the evolution of a deformation may be conceptualised as integrating the following equation:

$$\frac{d\varphi}{dt} = \mathbf{v}_t(\varphi)$$

Geodesic distances (from zero) are then measured by:

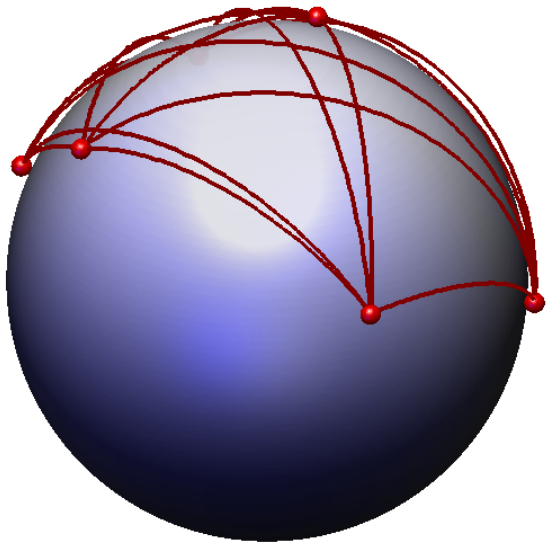
$$d = \int_{t=0}^{1} ||\mathbf{L}\mathbf{v}_t|| dt$$

Metric distances from large deformations



Miller et al. "Collaborative computational anatomy: an MRI morphometry study of the human brain via diffeomorphic metric mapping." Human Brain Mapping 30(7):2132–2141 (2009).

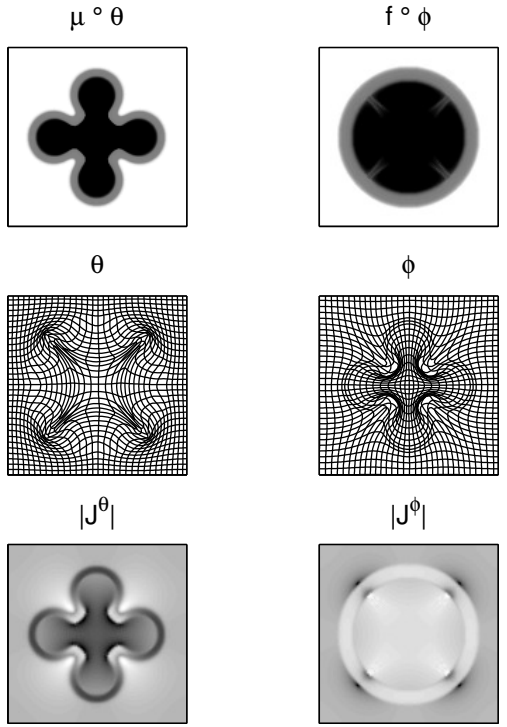
Metric distances from large deformations



Miller et al. "Collaborative computational anatomy: an MRI morphometry study of the human brain via diffeomorphic metric mapping." Human Brain Mapping 30(7):2132–2141 (2009).

Image Registration

- Image registration finds shortest distance between images.
- Often formulated to minimise the sum of two terms:
  - Distance between the image intensities.
  - Distance of the deformation from the identity.
- The sum of these gives a distance.



LDDMM

Large Deformation Diffeomorphic Metric Mapping is an image registration algorithm that minimises the following:

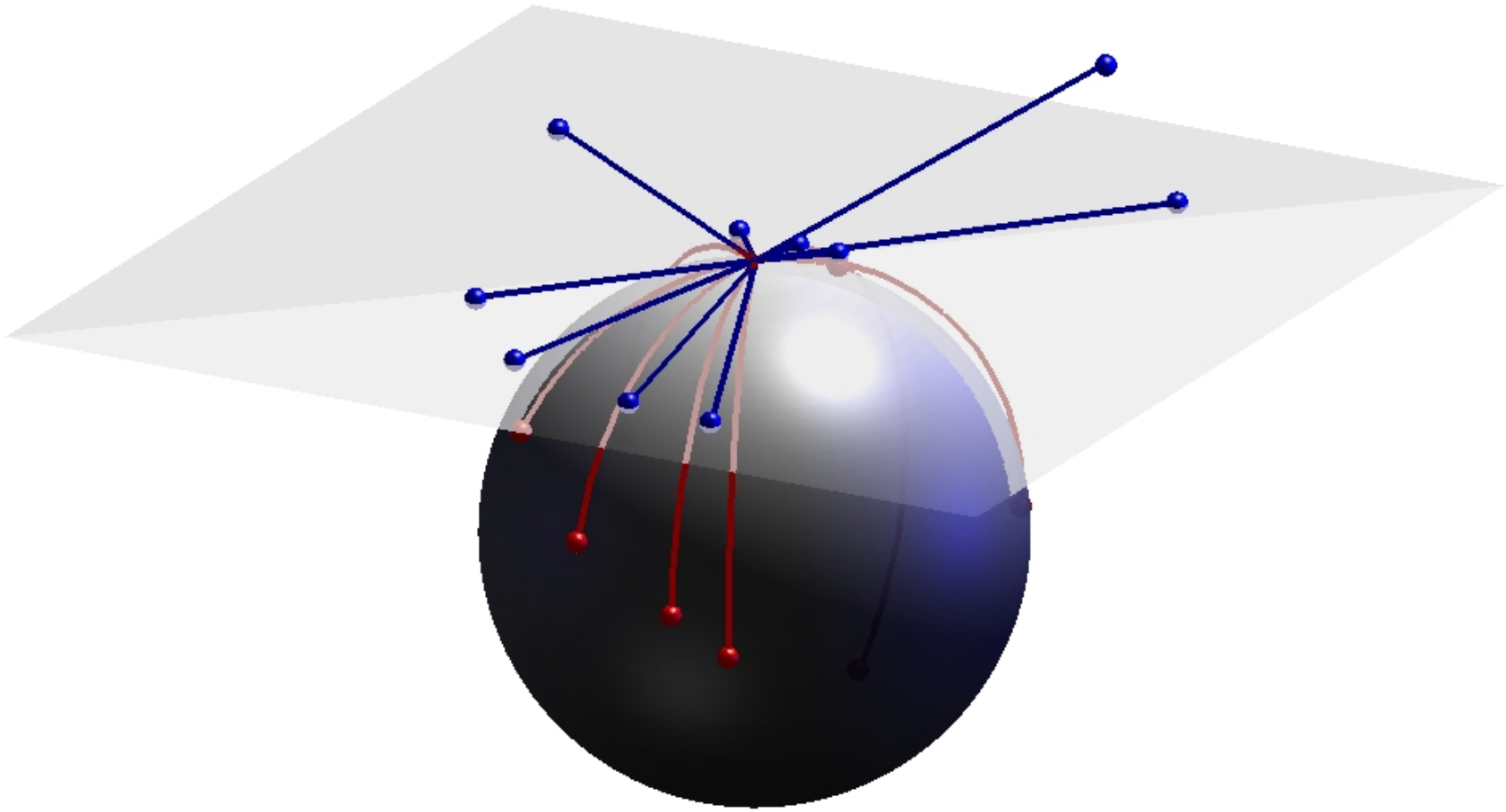
$$\mathcal{E} = \frac{1}{2}\int_{t=0}^{1} ||\mathbf{L}\mathbf{v}_t||^2 dt + \frac{1}{2\sigma^2}||f - \mu\left(\varphi_1^{-1}\right)||^2$$

$$\text{where } \varphi_0 = \text{id}, \ \frac{d\varphi}{dt} = \mathbf{v}_t\left(\varphi_t\right)$$

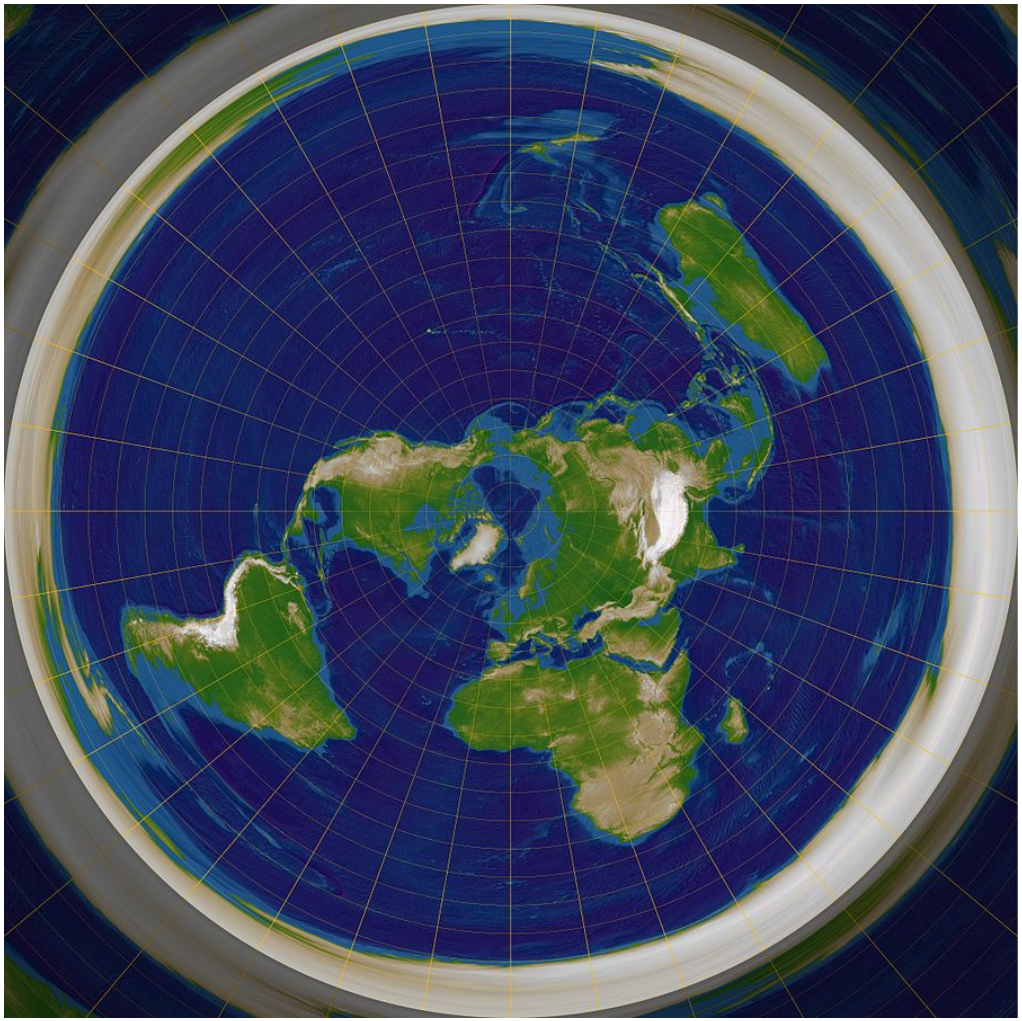
First term is a squared deformation distance measure.
Second term is the squared difference between images.
The objective is to estimate a series of velocity fields $(\mathbf{v}_t)$.

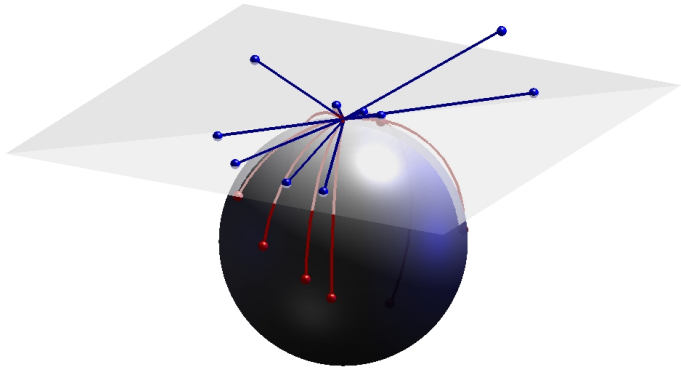
Beg, MF, Miller, MI, Trouvé, A & Younes, L. *Computing large deformation metric mappings via geodesic flows of diffeomorphisms*. International Journal of Computer Vision 61(2):139–157 (2005).

Linear approximations to nonlinear problems



Exponential map



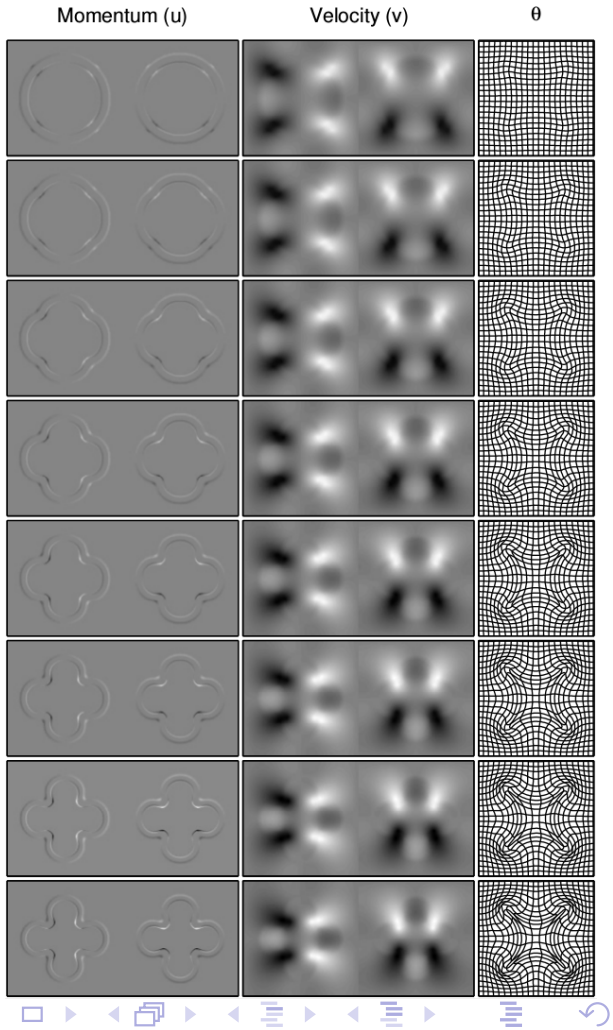


"Azimuthal Equidistant N90" by RokerHRO - Own work. Licensed under CC BY-SA 3.0 via Wikimedia Commons - http://commons.wikimedia.org/wiki/File:Azimuthal_Equidistant_N90.jpg

LDDMM via "geodesic shooting"

In practice, we just need to estimate an initial velocity ($\mathbf{v}_0$), from which we compute the initial momentum by $\mathbf{u}_0 = \mathbf{L}^\dagger \mathbf{L} \mathbf{v}_0$.
We set the deformation at time 0 to an identity transform ($\varphi_0 = id$), and then evolve the following dynamical system for unit time:

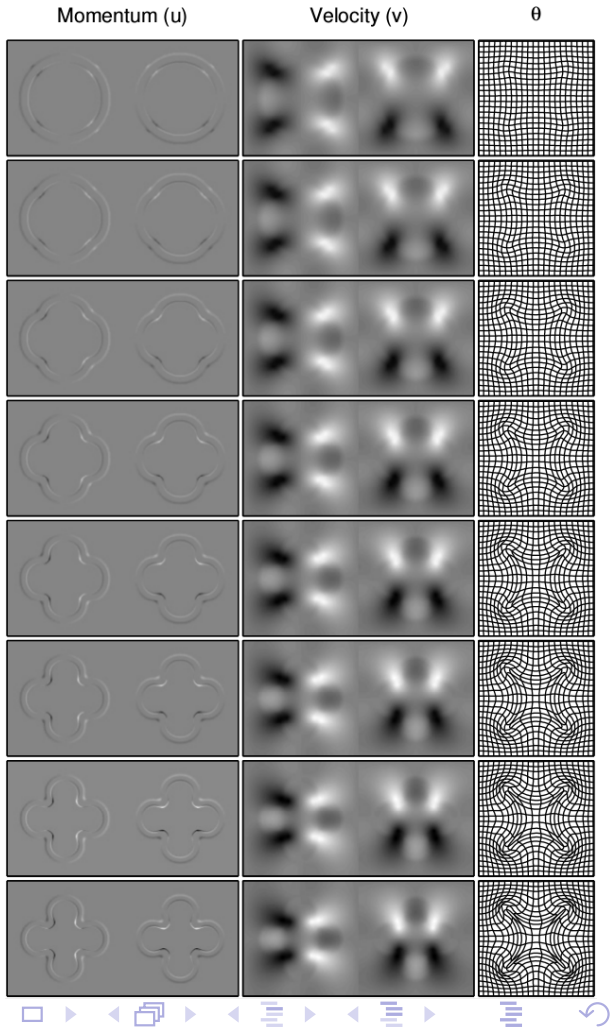
$$\begin{aligned}
\mathbf{u}_t &= \det|\mathbf{D}\varphi_t^{-1}|(\mathbf{D}\varphi_t^{-1})^T(\mathbf{u}_0 \circ \varphi_t^{-1}) \\
\mathbf{v}_t &= \left(\mathbf{L}^\dagger \mathbf{L}\right)^{-1} \mathbf{u}_t \\
\frac{d\varphi}{dt} &= \mathbf{v}_t(\varphi_t)
\end{aligned}$$



LDDMM via "geodesic shooting"

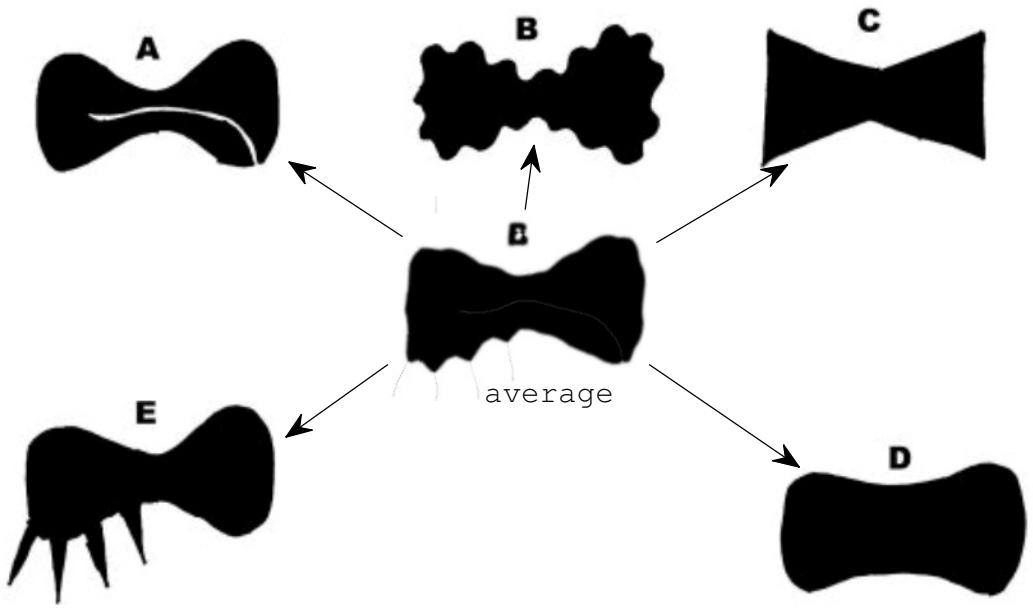
The final deformation (φ_1) is a type of exponential of the initial velocity ($\mathbf{v}_0$).

Exponential map (Riemannian geometry). (2015, January 13). In Wikipedia, The Free Encyclopedia. Retrieved 18:04, March 31, 2015, from `http://en.wikipedia.org/w/index.php?title=Exponential_map_(Riemannian_geometry)&oldid=642372186`



"Groupwise registration"

Minimising distortions by centering around the mean.

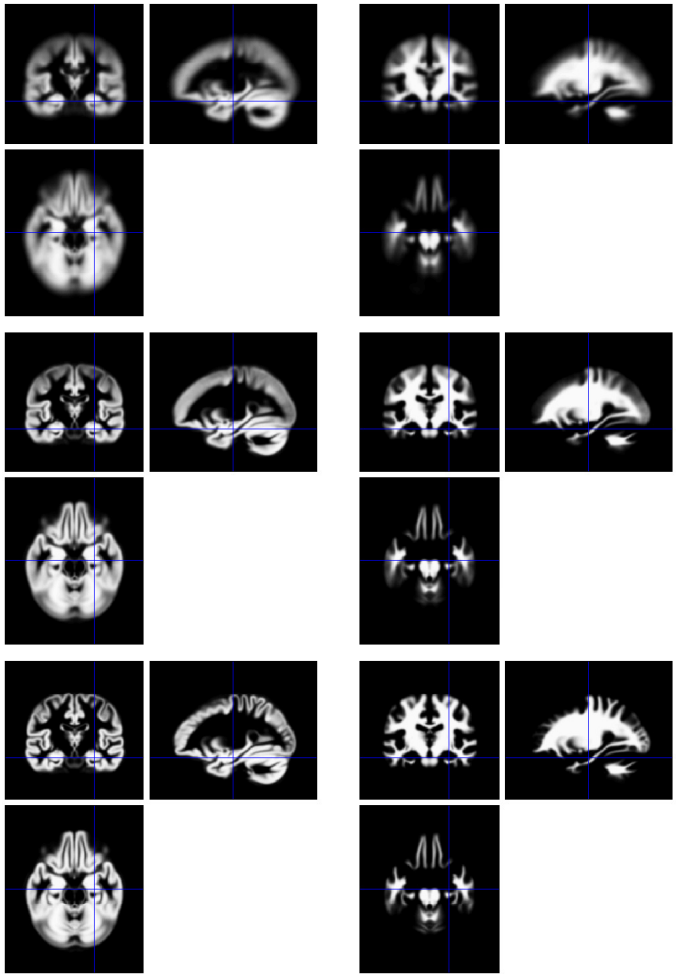


"Groupwise registration"

"Groupwise registration"

Ignoring the many technical details, the procedure involves alternating between:

- Create the mean of aligned images.
- Align all images to be slightly closer to the mean.



Kernel matrix

Construction of kernel matrix accounts for the regularisation used by the image registration:

$$k(\mathbf{v}_i, \mathbf{v}_j) = \langle \mathbf{L}^\dagger \mathbf{L} \mathbf{v}_i, \mathbf{v}_j \rangle$$
$$= \langle \mathbf{L}\mathbf{v}_i, \mathbf{L}\mathbf{v}_j \rangle$$

Wang, Lei, et al. "Large deformation diffeomorphism and momentum based hippocampal shape discrimination in dementia of the Alzheimer type." Medical Imaging, IEEE Transactions on 26(4):462–470 (2007).

"Scalar momentum"

At the solution, gradients of the LDDMM objective function should vanish:

$$\mathbf{L}^\dagger \mathbf{L}\mathbf{v}_0 + \frac{1}{\sigma^2} \det |\mathbf{D}\boldsymbol{\varphi}_1| (f \circ \boldsymbol{\varphi}_1 - \mu)(\nabla \mu) = 0$$

Re-expressiong this, we see that the initial velocity (and momentum) is given by:

$$\mathbf{L}^\dagger \mathbf{L}\mathbf{v}_0 = \mathbf{u}_0 = \frac{1}{\sigma^2} (\nabla \mu) \det |\mathbf{D}\boldsymbol{\varphi}_1| (\mu - f \circ \boldsymbol{\varphi}_1)$$

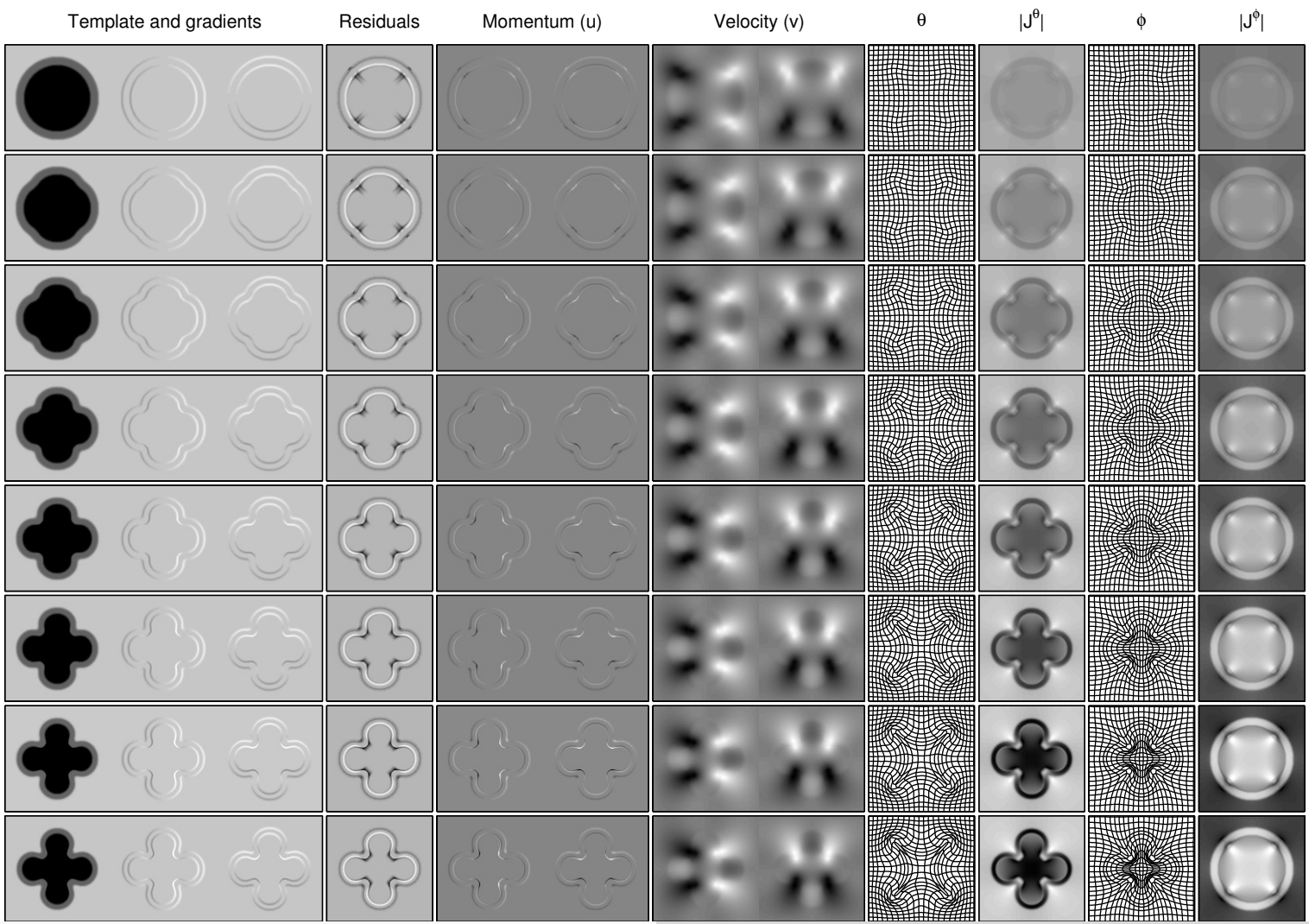
"Scalar Momentum"

$$\mathbf{u}_0 = \frac{1}{\sigma^2}(\nabla\mu)\det|\mathbf{D}\varphi_1|(\mu - f\circ\varphi_1)$$

If a population of subjects are all aligned with the same template image, $\frac{1}{\sigma^2}(\nabla\mu)$ will be the same for all subjects. Deviations from the template are encoded by the "*scalar momentum*", $\det|\mathbf{D}\varphi_1|(\mu - f\circ\varphi_1)$. This is a scalar field, and in principle is all that is needed (along with the template) to reconstruct the original images.

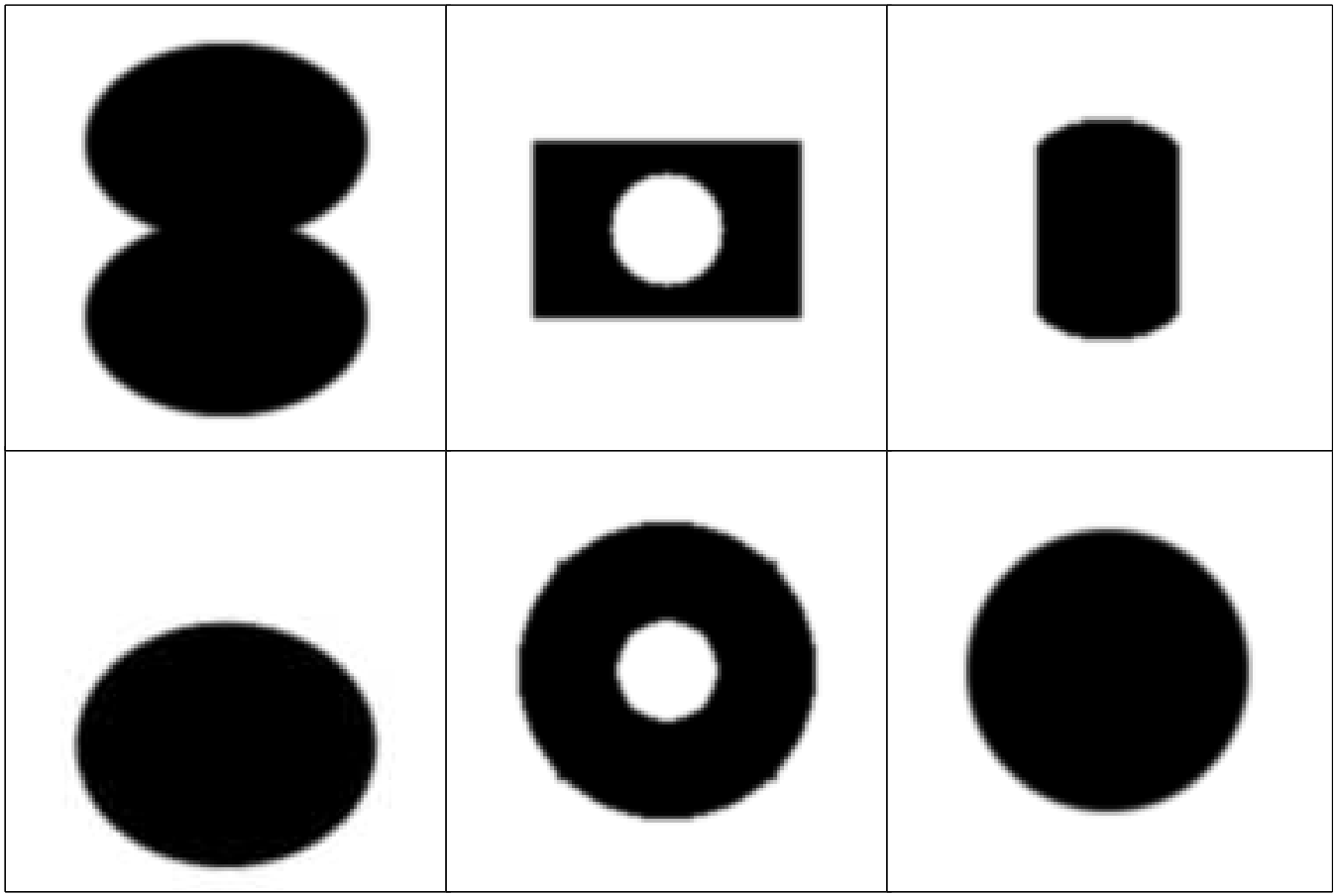
Miller et al. "Collaborative computational anatomy: an MRI morphometry study of the human brain via diffeomorphic metric mapping." Human Brain Mapping 30(7):2132–2141 (2009).
Singh, Fletcher, Preston, Ha, King, Marron, Wiener & Joshi (2010). *Multivariate Statistical Analysis of Deformation Momenta Relating Anatomical Shape to Neuropsychological Measures*. T. Jiang et al. (Eds.): MICCAI 2010, Part III, LNCS 6363, pp. 529–537, 2010.

Evolution


Template and gradients
Residuals
Momentum (u)
Velocity (v)
θ
|J^θ|
ϕ
|J^ϕ|

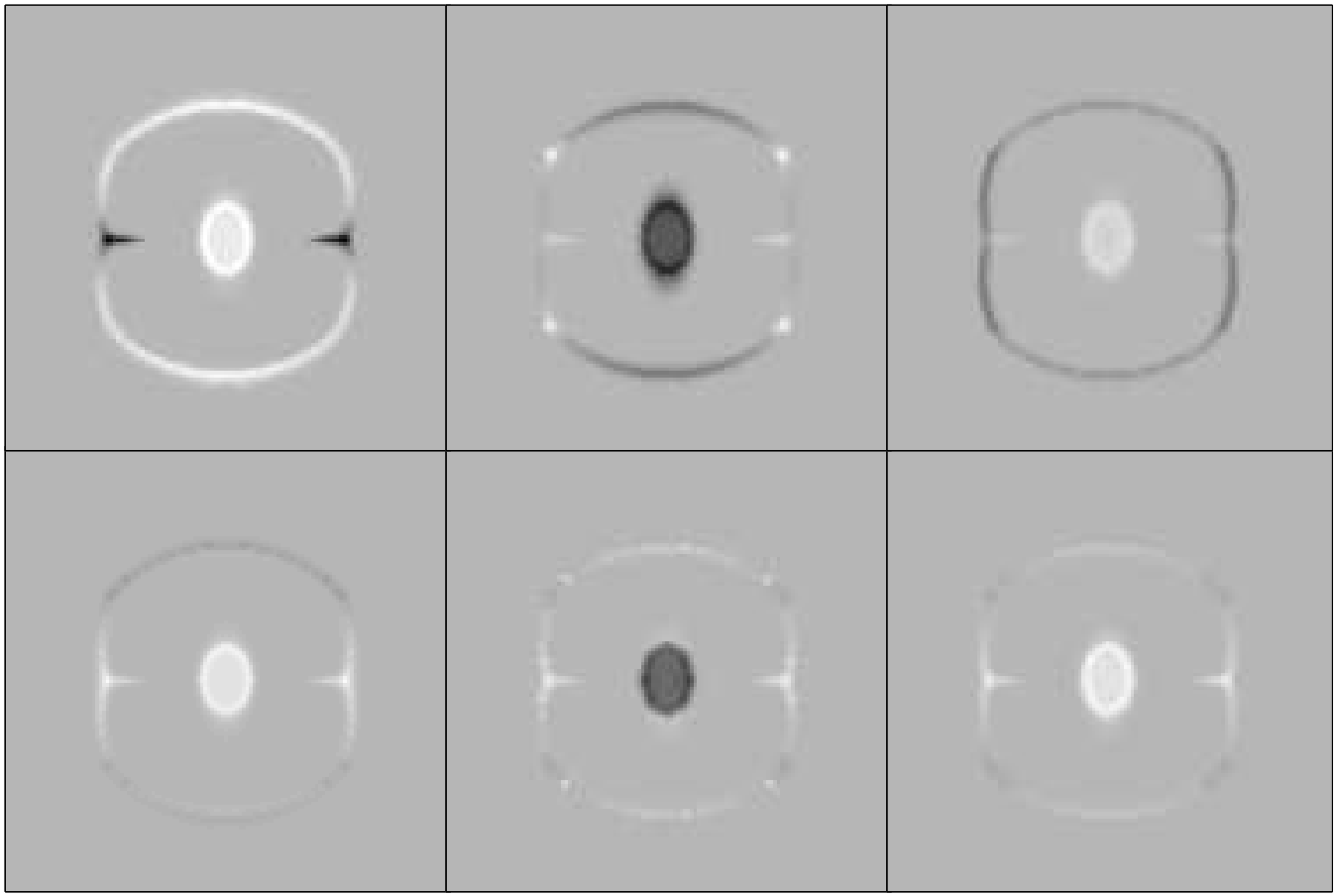
Example Images

Some example images.



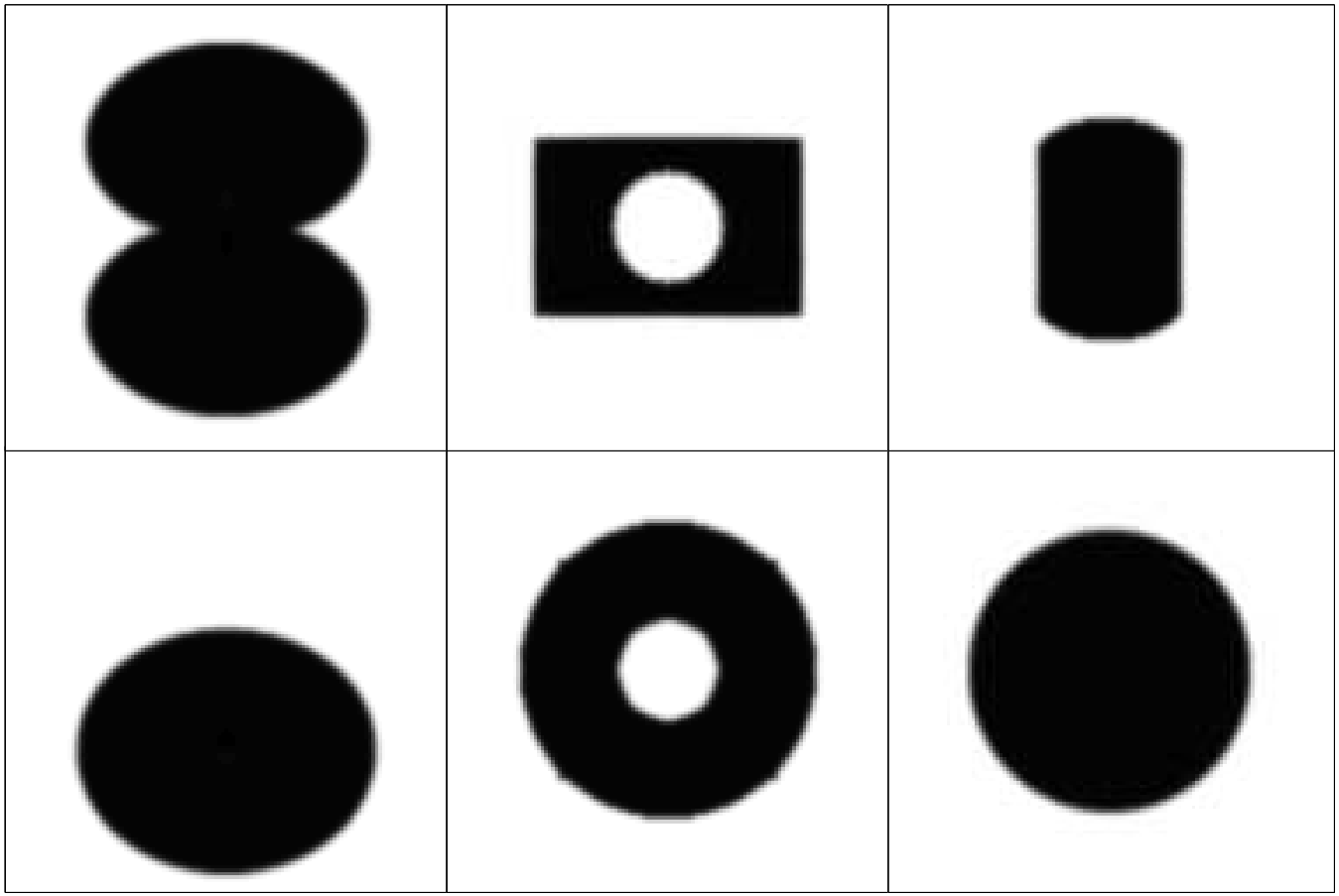
Scalar Momentum

Scalar momenta after aligning to a common template.



Reconstructed Images

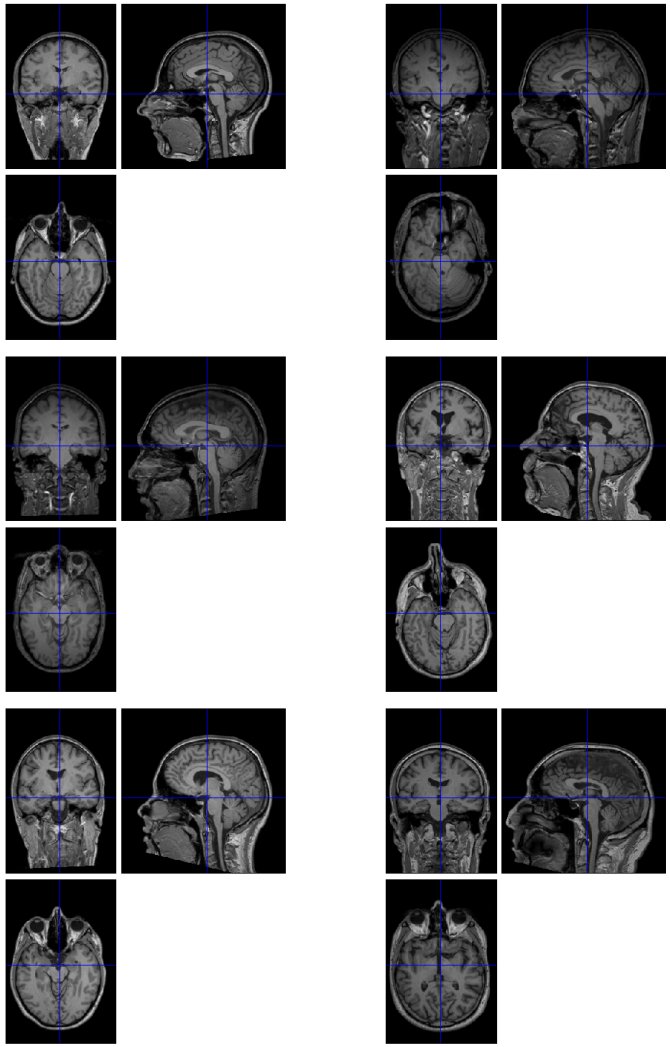
Images reconstructed from scalar momenta (and template).



Real data

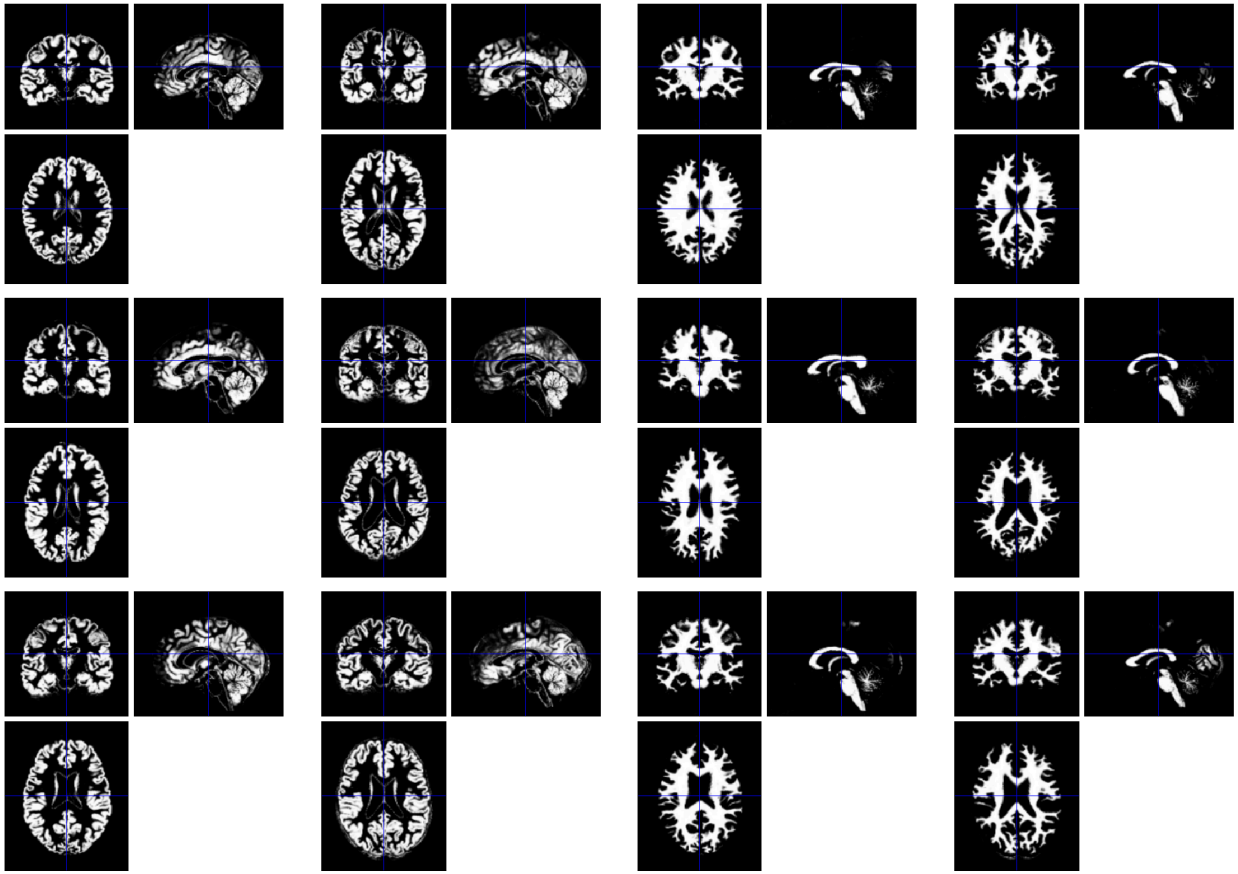
Used 550 T1w brain MRI from IXI (**I**nformation e**X**traction from **I**mages) dataset.
`http://www.brain-development.org/`
Data from three different hospitals in London:

- Hammersmith Hospital using a Philips 3T system
- Guy's Hospital using a Philips 1.5T system
- Institute of Psychiatry using a GE 1.5T system



Grey and White Matter

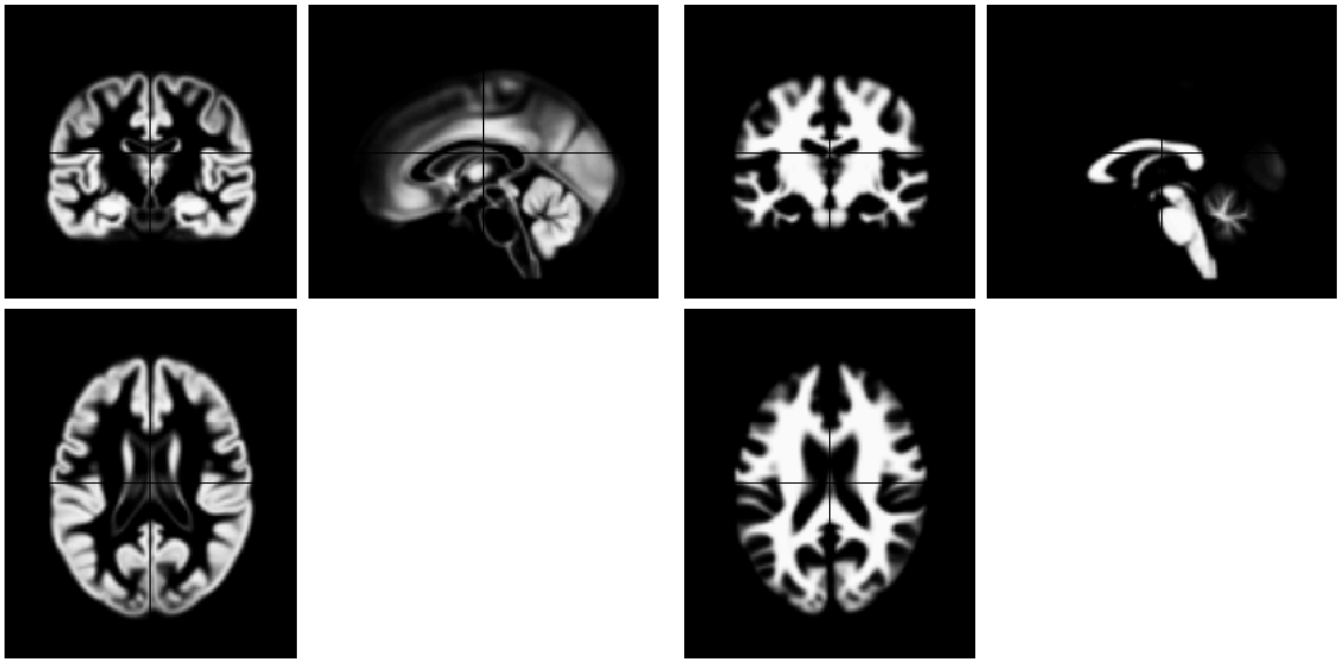
Segmented into GM and WM. Approximately aligned via rigid-body.



Ashburner, J & Friston, KJ. *Unified segmentation*. NeuroImage 26(3):839–851 (2005).

Diffeomorphic Alignment

All GM and WM were diffeomorphically aligned to their common average-shaped template.

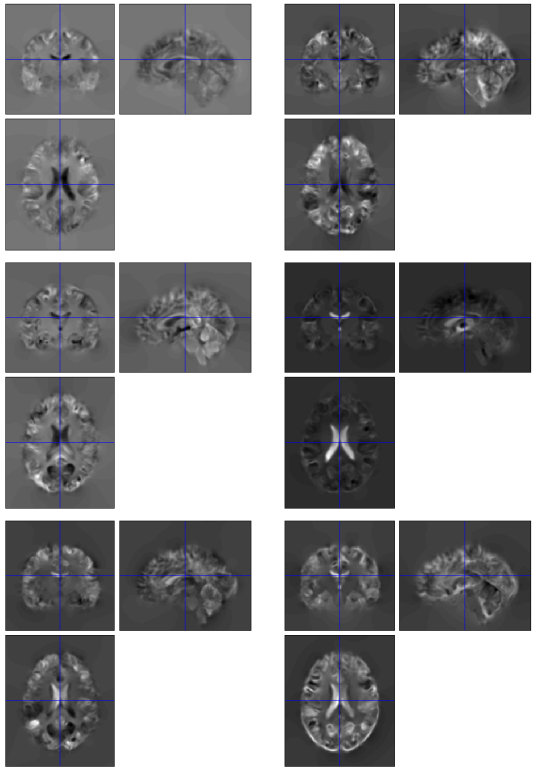


Ashburner, J & Friston, KJ. *Diffeomorphic registration using geodesic shooting and Gauss-Newton optimisation*. NeuroImage 55(3):954–967 (2011).
Ashburner, J & Friston, KJ. *Computing average shaped tissue probability templates*. NeuroImage 45(2):333–341 (2009).

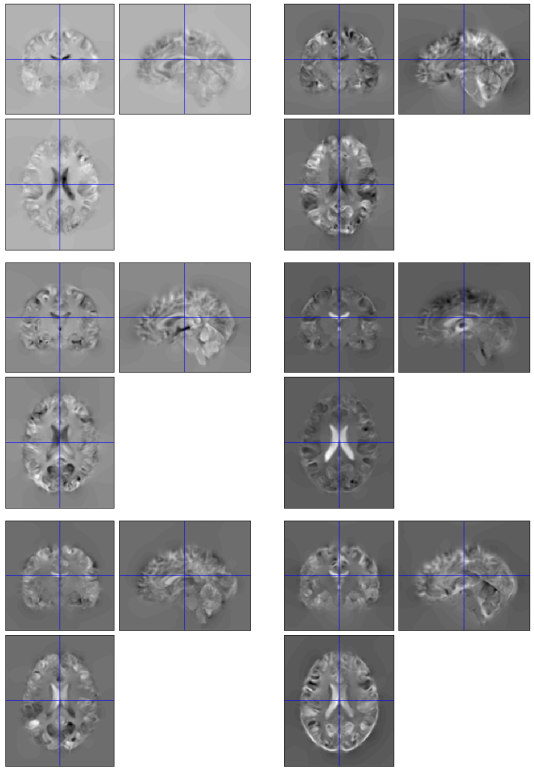
Volumetric Features

A number of features were used for pattern recognition.
Firstly, two features relating to relative volumes.
Initial velocity divergence is similar to logarithms of Jacobian determinants.

Jacobian Determinants

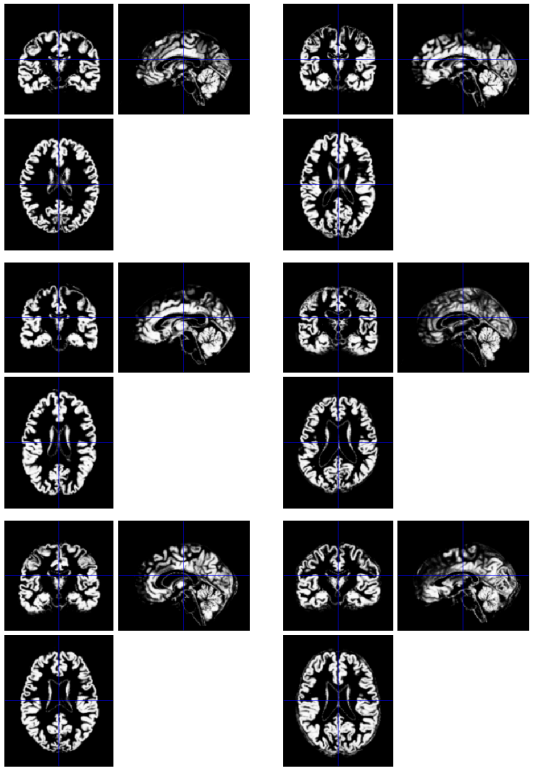


Initial Velocity Divergence

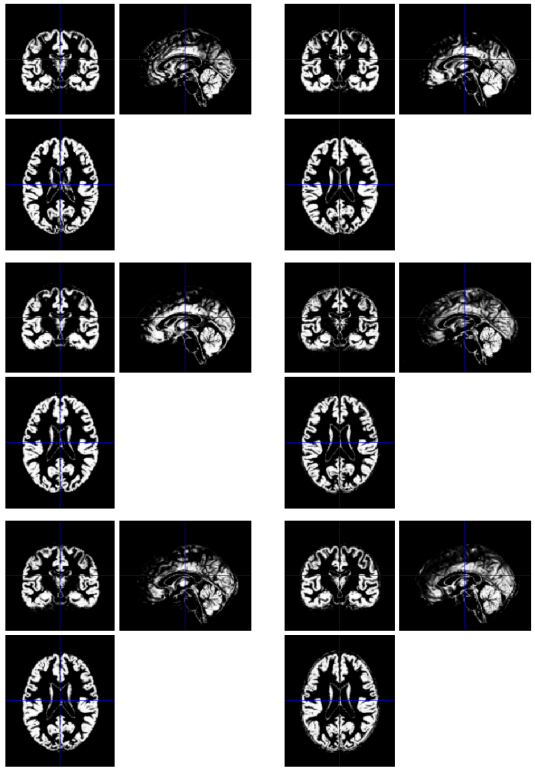


Grey Matter Features

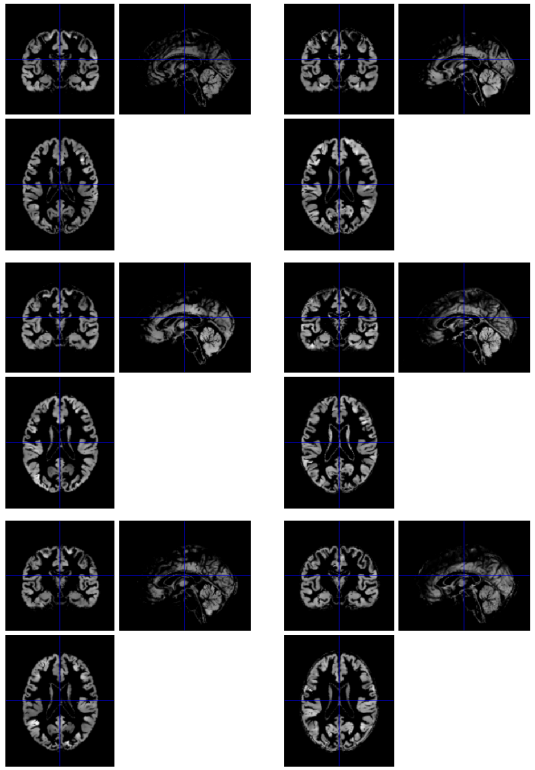
Rigidly Registered GM



Nonlinearly Registered GM



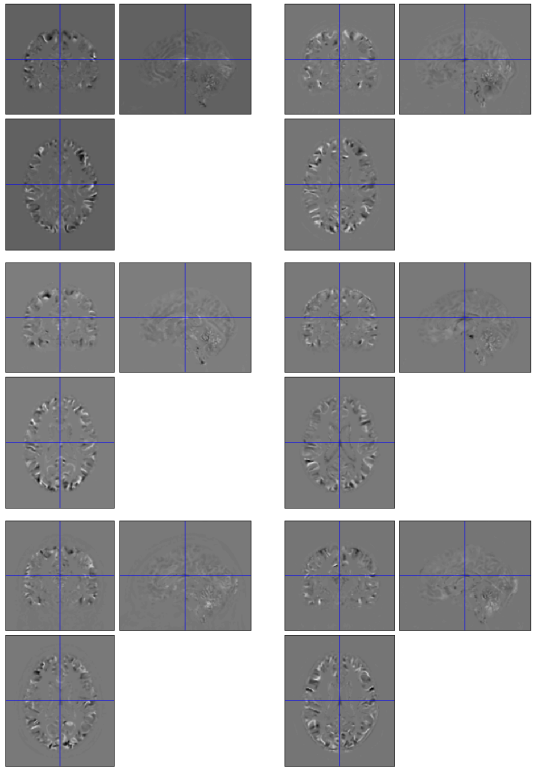
Registered and Jacobian Scaled GM



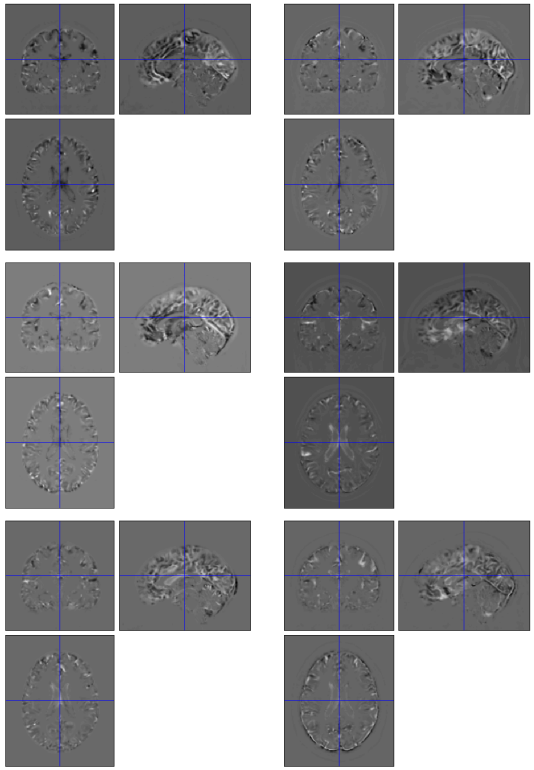
"Scalar Momentum" Features

"Scalar momentum" actually has two components because GM was matched with GM and WM was matched with WM.

First Momentum Component

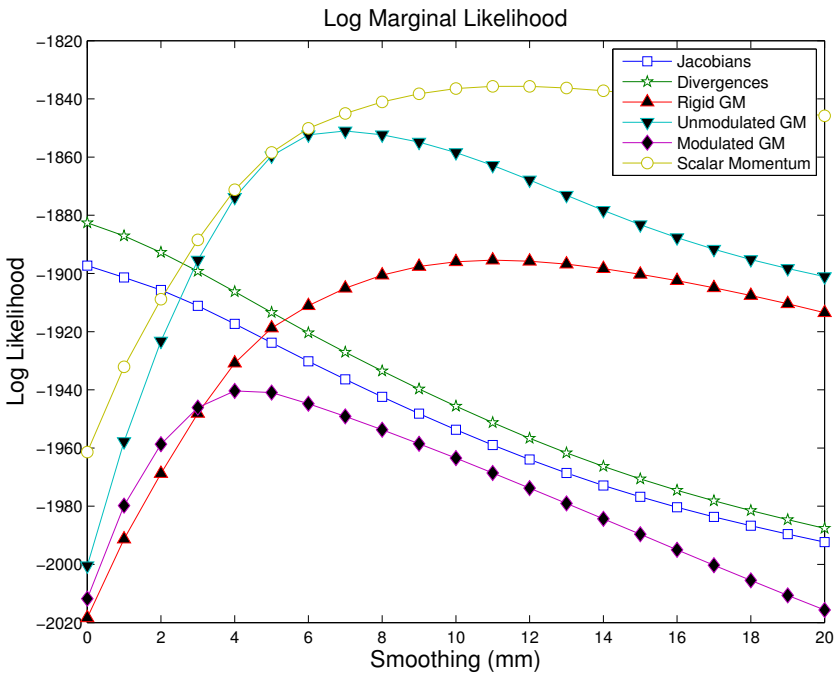


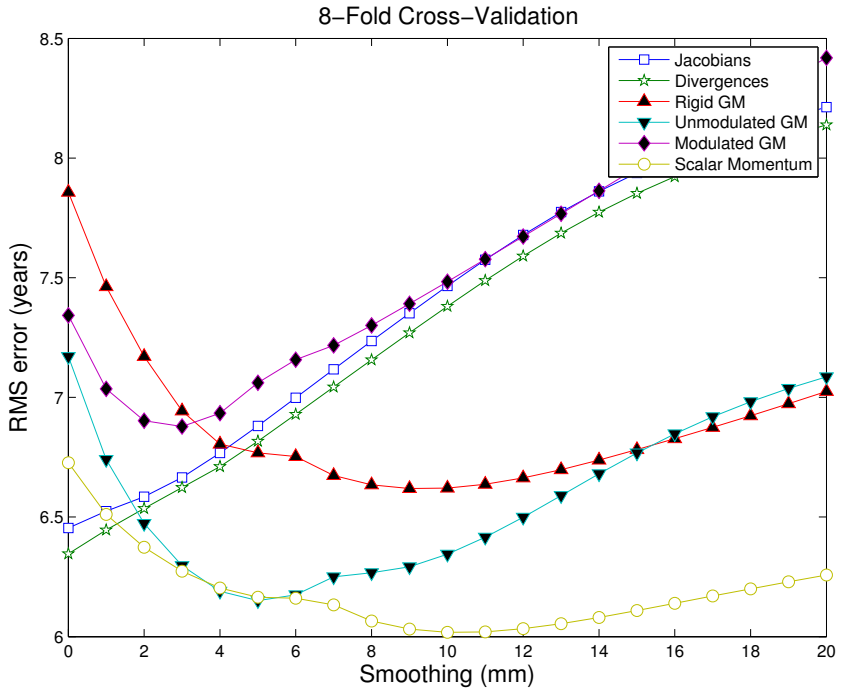
Second Momentum Component



Age Regression

Linear Gaussian Process Regression to predict subject ages.

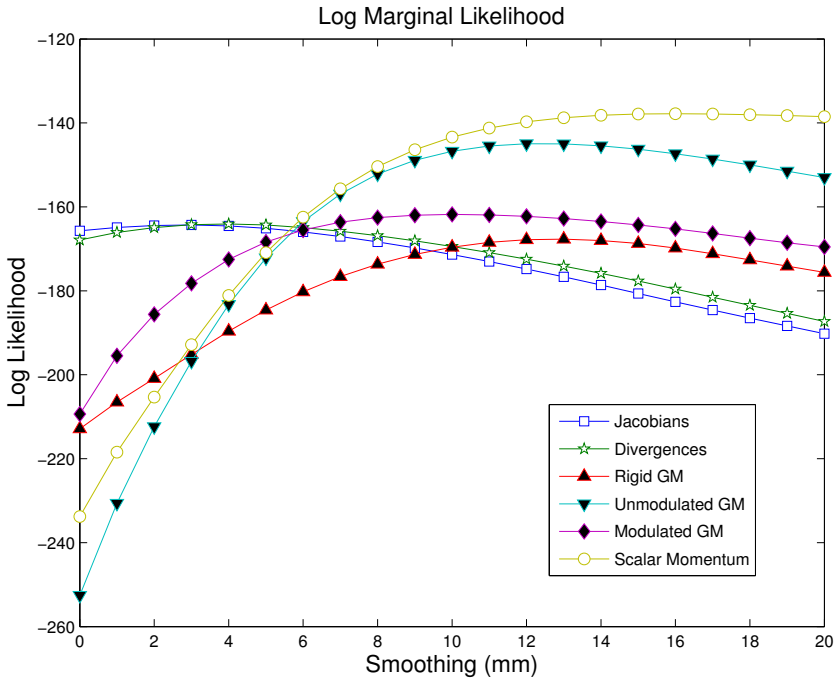


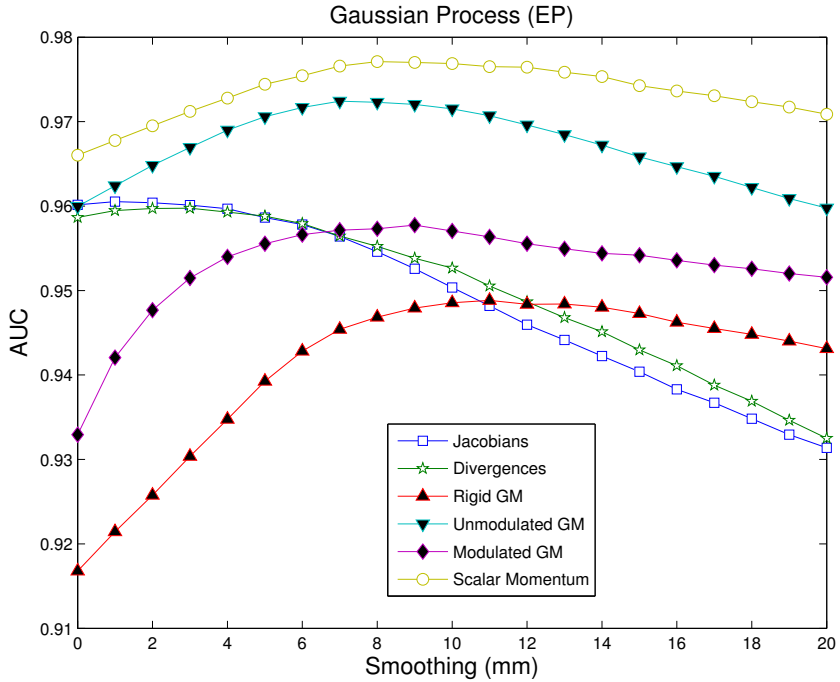


Rasmussen, CE & Williams, CKI. *Gaussian processes for machine learning*. Springer (2006).

Sex Classification

Linear Gaussian Process Classification (EP) to predict sexes.

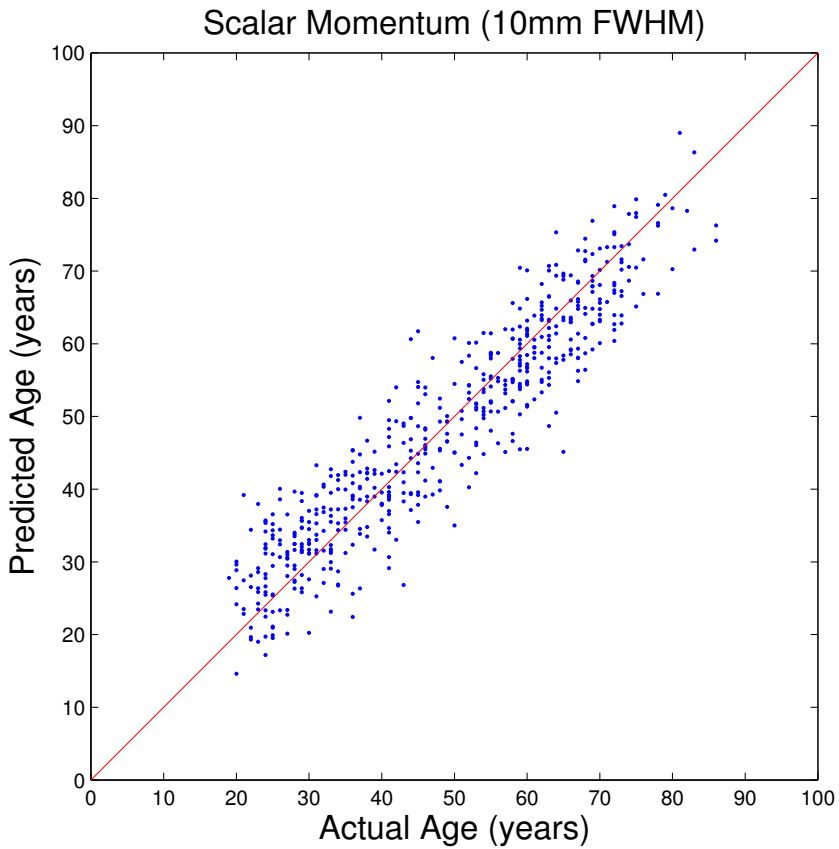




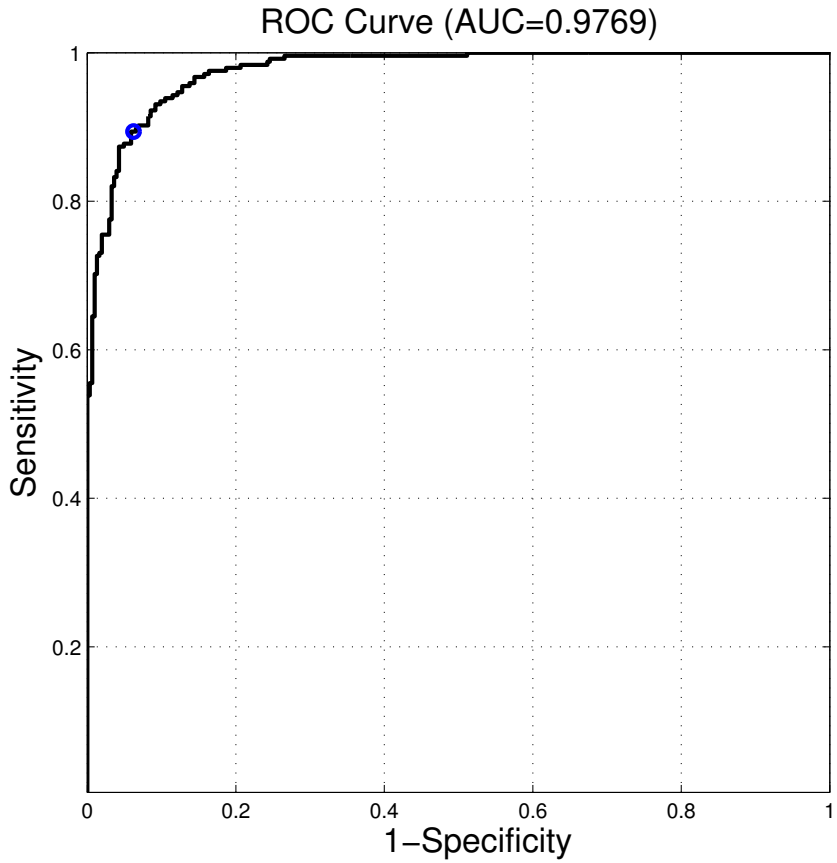
Rasmussen, CE & Williams, CKI. *Gaussian processes for machine learning*. Springer (2006).

Predictive Accuracies

Age



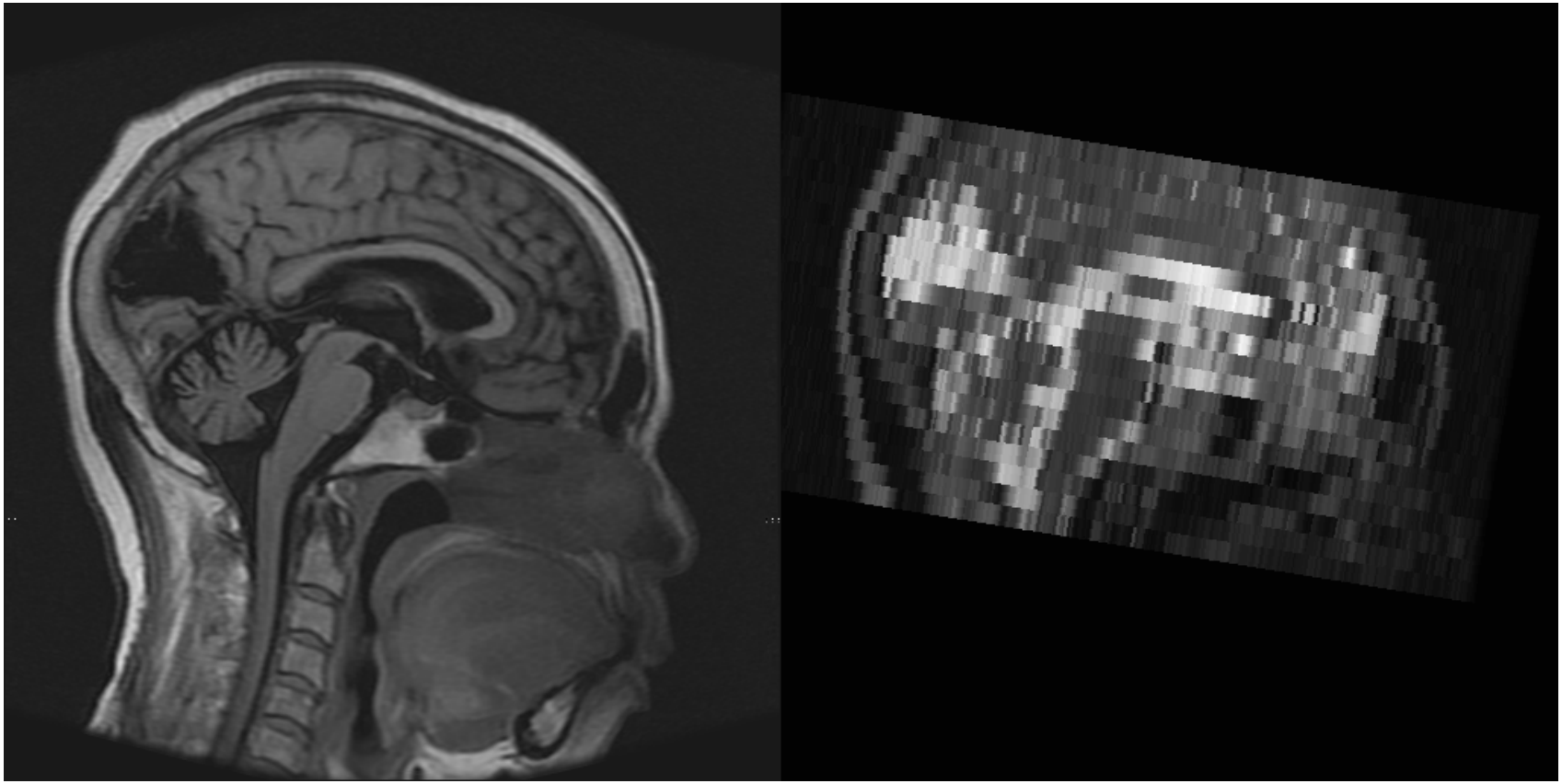
Sex



Conclusions

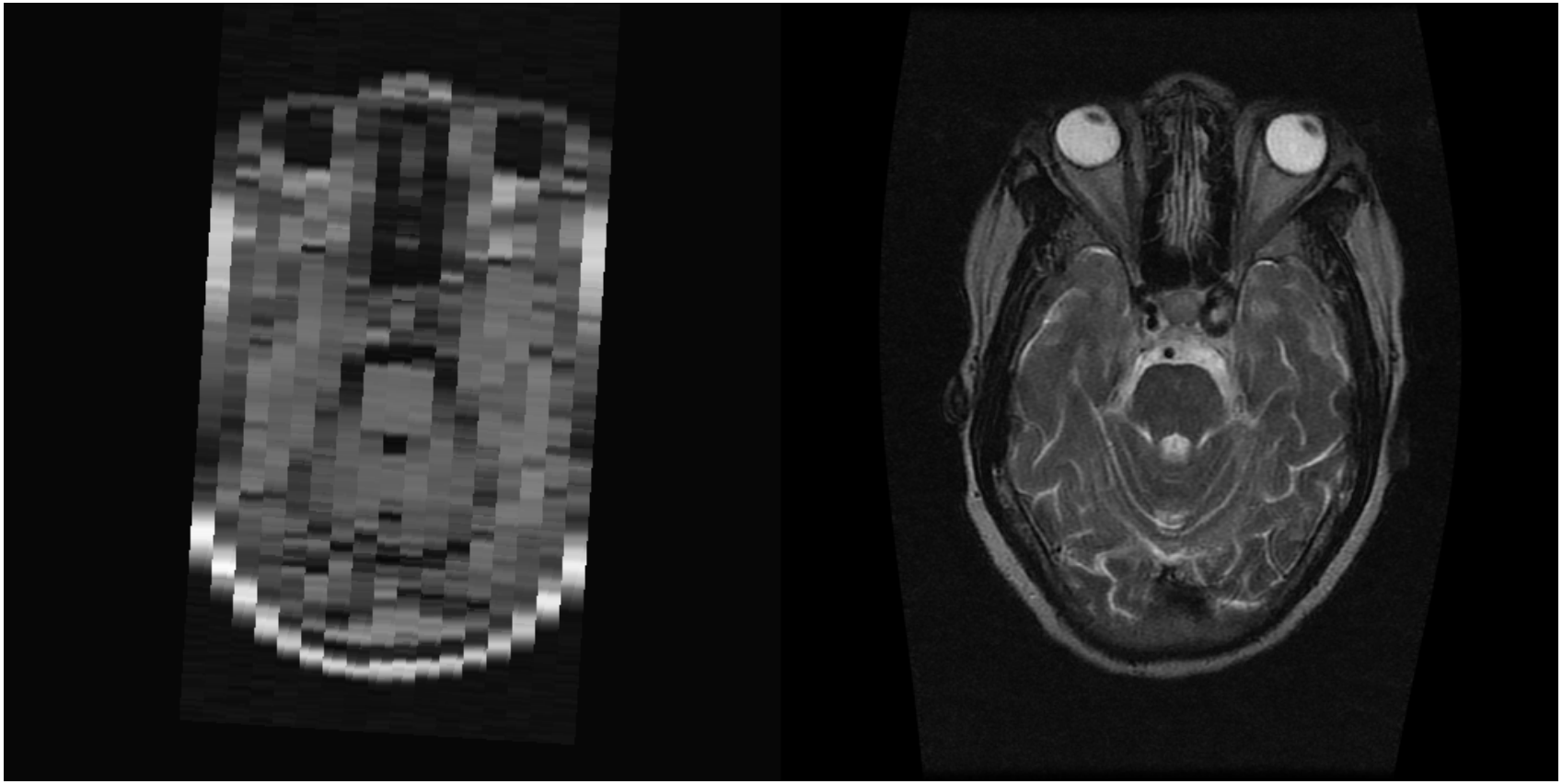
- Scalar momentum (with about 10mm smoothing) appears to be a useful feature set.
- Jacobian-scaled warped GM is surprisingly poor.
- Amount of spatial smoothing makes a big difference.
- Further dependencies on the details of the registration still need exploring.

Mining hospital data



High resolution in-plane. Very thick slices.
Multiple image contrasts/orientations.

Mining hospital data



High resolution in-plane. Very thick slices.
Multiple image contrasts/orientations.

Data sharing

Increasing pressure from funding bodies to share data.

- Push for reproducible research.
- Allows more discovery science.

... conflicts with personal privacy.

`http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/`
`http://grants.nih.gov/grants/policy/data_sharing/`

Bigger data

UK Biobank Imaging UK Biobank is a *long-term prospective epidemiological study* that has already collected genetics, blood samples, lifestyle information and other data from a cohort of 500,000 subjects, to be followed clinically over coming decades. The UK Biobank Imaging Extension, which aims to bring back *100,000 of the cohort for multimodal neuroimaging* and cardiac MRI (amongst other measures), has just been given the go-ahead. This will be by far the largest neuro/cardiac imaging study carried out to date, and will add very rich phenotyping to the overall Biobank project.

Better use of diagnostic information

Information is lost when creating binary labels.

1. **Normal subjects**: MMSE scores between 24-30 (inclusive), a CDR of 0, non-depressed, non MCI, and nondemented. The age range of normal subjects will be roughly matched to that of MCI and AD subjects. Therefore, there should be minimal enrollment of normals under the age of 70.
2. **MCI subjects**: MMSE scores between 24-30 (inclusive), a memory complaint, have objective memory loss measured by education adjusted scores on Wechsler Memory Scale Logical Memory II, a CDR of 0.5, absence of significant levels of impairment in other cognitive domains, essentially preserved activities of daily living, and an absence of dementia.
3. **Mild AD**: MMSE scores between 20-26 (inclusive), CDR of 0.5 or 1.0, and meets NINCDS/ADRDA criteria for probable AD.

`http://www.adni-info.org/scientists/ADNIGrant/ProtocolSummary.aspx`

Uncertain labels

...all clinically diagnosed AD patients will not have AD pathology, and up to 30% of cognitively normal subjects will have a pathologic diagnosis of AD at autopsy.

Vemuri et al. "Antemortem MRI based STructural Abnormality iNDex (STAND)-scores correlate with postmortem Braak neurofibrillary tangle stage". NeuroImage 42(2):559–567 (2008).

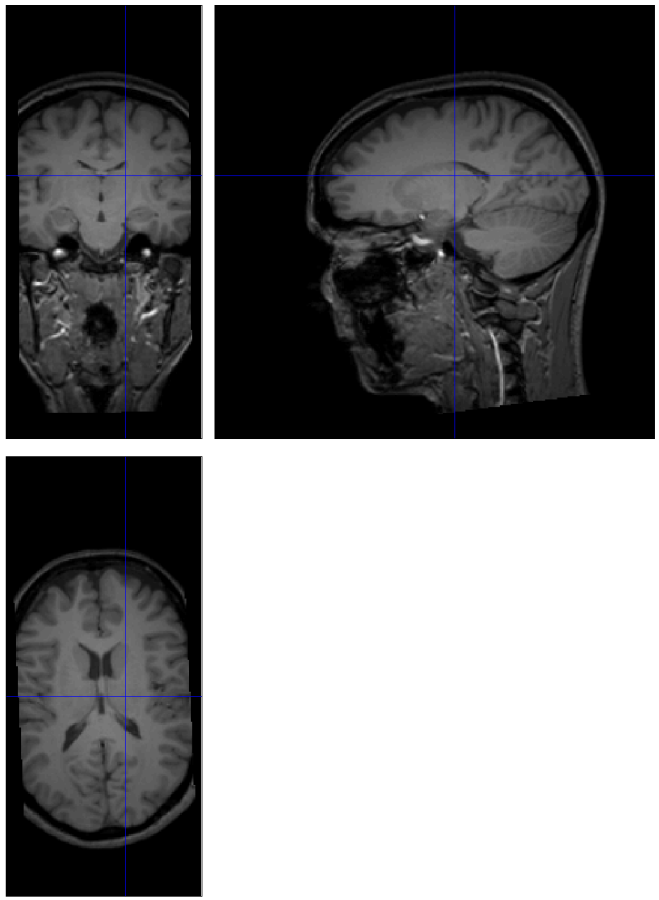
Model-based dimensionality reduction

Nonlinear dimensionality reduction techniques related to PCA:

- **Principal geodesic analysis** combines PCA with image registration to maximise the amount of signal explained.
  Zhang, Miaomiao, and P. Thomas Fletcher. "Bayesian Principal Geodesic Analysis in Diffeomorphic Image Registration." Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014. Springer International Publishing, 2014. 121-128.
- **Non-negative matrix factorization** decomposes data into matrices that are non-negative.
  Lee, Daniel D., and H. Sebastian Seung. "Learning the parts of objects by non-negative matrix factorization." Nature 401.6755 (1999): 788-791.
  Sotiras, Aristeidis, Susan M. Resnick, and Christos Davatzikos. "Finding imaging patterns of structural covariance via Non-Negative Matrix Factorization." NeuroImage 108 (2015): 1-16.
- **Generalized principal components** incorporates a link function into principal component analysis.
  Collins, Michael, Sanjoy Dasgupta, and Robert E. Schapire. "A generalization of principal components analysis to the exponential family." Advances in neural information processing systems. 2001.
  Mohamed, Shakir, Zoubin Ghahramani, and Katherine A. Heller. "Bayesian exponential family PCA." Advances in Neural Information Processing Systems. 2009.

Missing data

- Brain images of different individuals have different fields of view.
- Currently, if test subject has smaller field of view than training data, the pattern recognition approch needs to be retrained.
- Improved generative modelling could be used to better deal with missing data.



Principled similarity measures

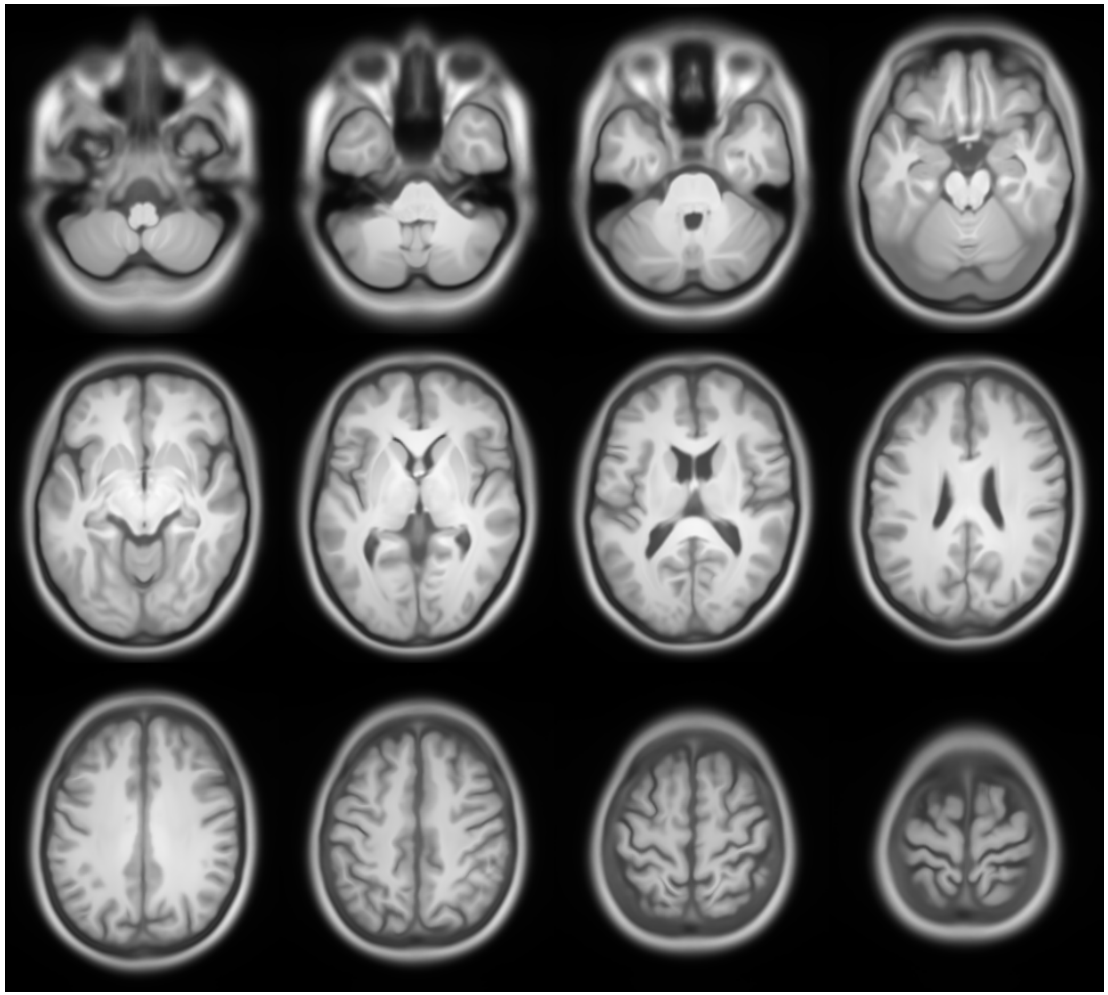
How do we best compute similarity from models used to "pre-process" data?

- **Fisher kernels** T. S. Jaakkola and D. Haussler. "Exploiting generative models in discriminative classifiers." In Kearns et al. [26], pages 487–493.

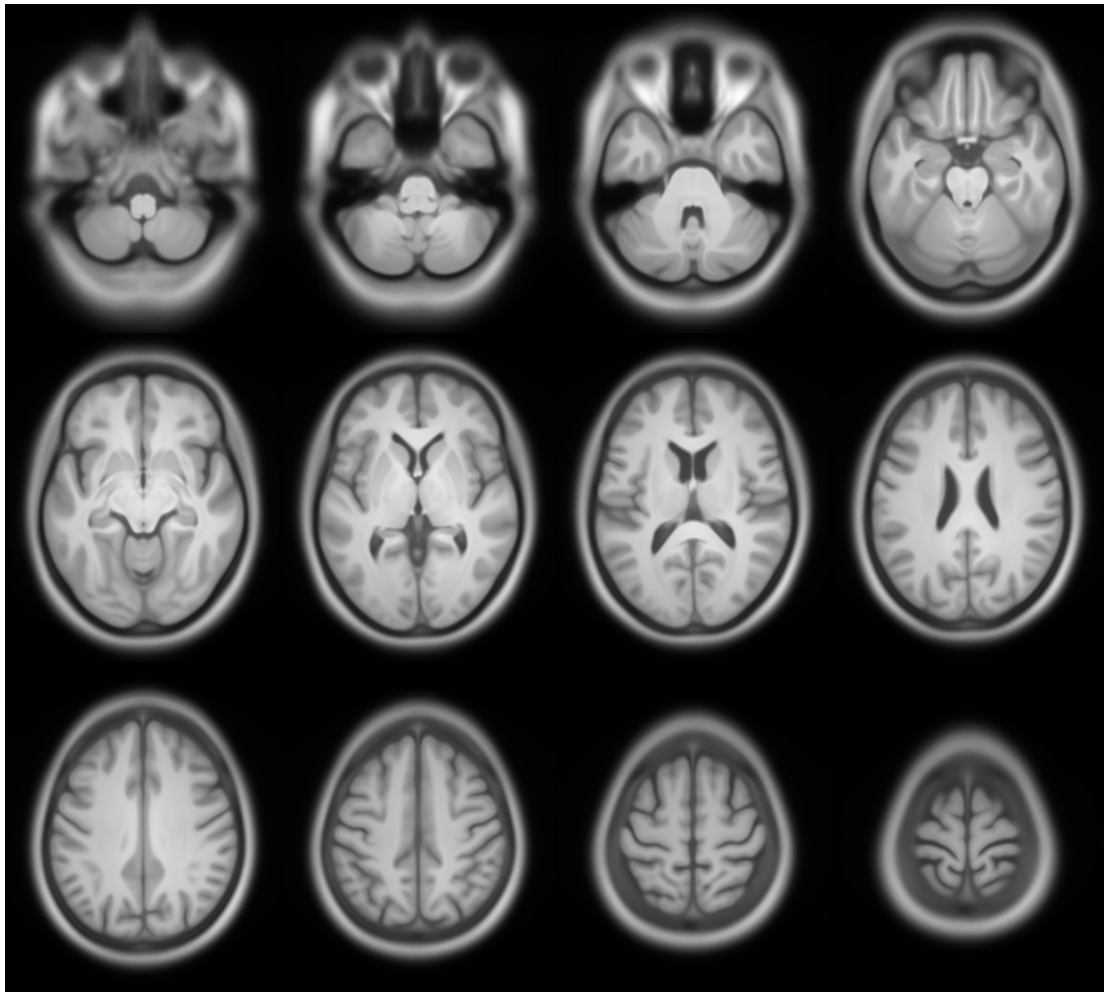
Visualising differences

Neuroimagers like blobs and summary tables.
How do we best understand whole-brain multivariate differences?
Here's my attempt.....

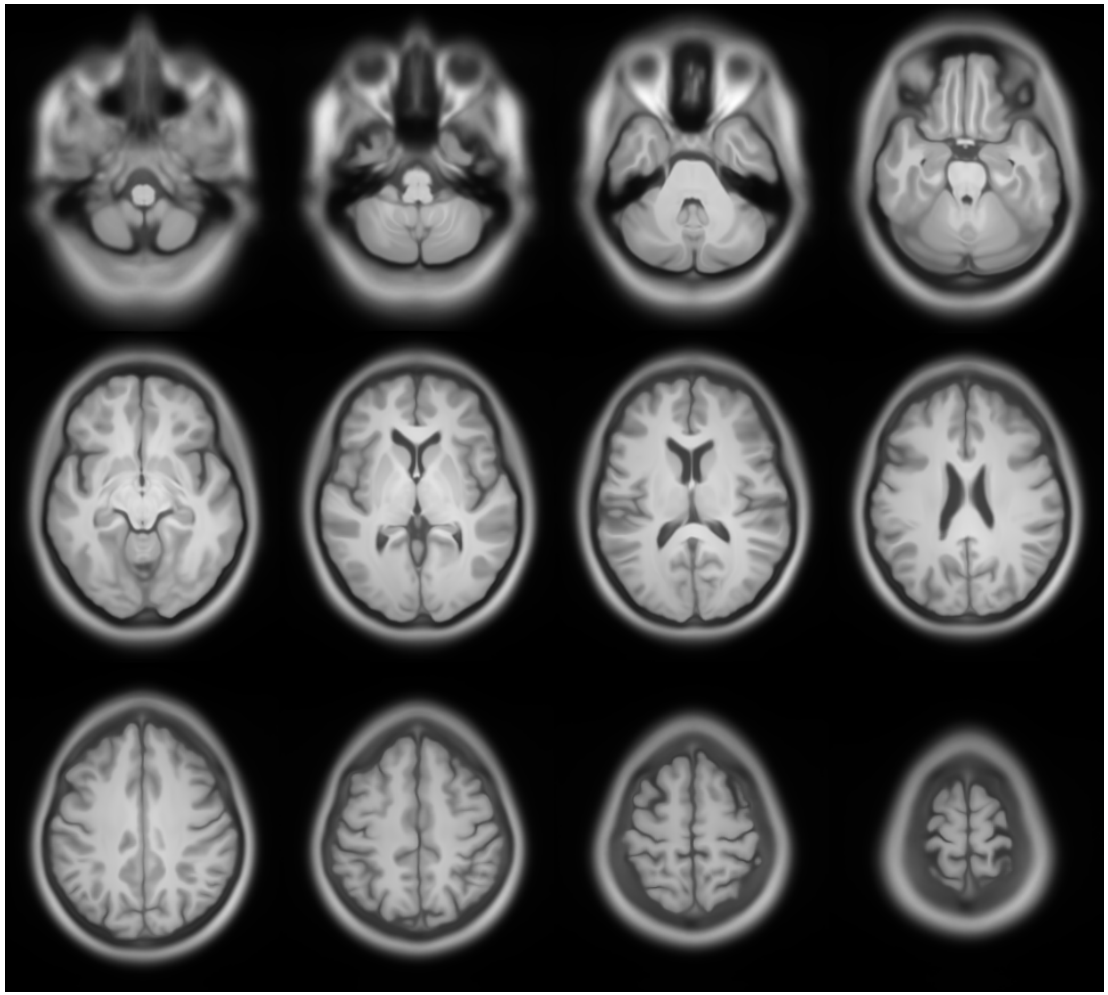
Exaggerated male brain



Average brain



Exaggerated female brain



More information

- These slides are available from `https://github.com/JohnAshburner/ISBI`
- For more information, see
  Ashburner, John, and Stefan Klöppel. "Multivariate models of inter-subject anatomical variability." Neuroimage 56.2 (2011): 422-439.