Generative Models for Medical Imaging

John Ashburner

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

1 Introduction
- Pipelines v Models
- Probability Theory
- Medical image computing

2 Segmentation

3 Diffeomorphic Registration

4 Longitudinal Registration

5 Dimensionality Reduction

Moravec's Paradox

Rodney Brooks explains that, according to early AI research, intelligence was "best characterized as the things that highly educated male scientists found challenging", such as chess, symbolic integration, proving mathematical theorems and solving complicated word algebra problems. "The things that children of four or five years could do effortlessly, such as visually distinguishing between a coffee cup and a chair, or walking around on two legs, or finding their way from their bedroom to the living room were not thought of as activities requiring intelligence."

Moravec's paradox. (2015, April 25). In Wikipedia, The Free Encyclopedia. Retrieved 14:46, June 17, 2015, from `https://en.wikipedia.org/w/index.php?title=Moravec%27s_paradox&oldid=659139375`

`https://xkcd.com`

Why Image Processing Seems Easy

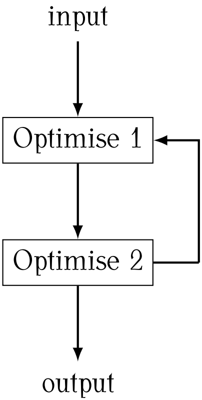
Neurons for visual processing take up 30% of the brain's cortex (as opposed to about 8 % for touch and 3 % for hearing).

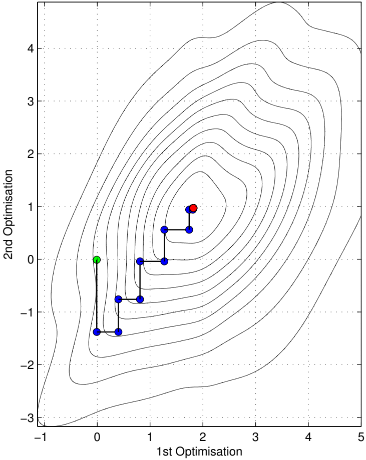
Pipelines

In software engineering, a pipeline consists of a chain of processing elements (processes, threads, coroutines, functions, etc.), arranged so that the output of each element is the input of the next

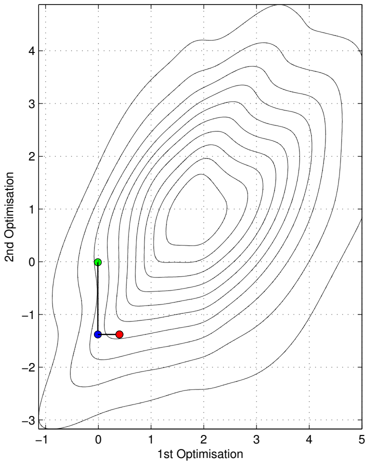
Pipeline (software). (2015, May 1). In Wikipedia, The Free Encyclopedia. Retrieved 16:50, June 17, 2015, from `https://en.wikipedia.org/w/index.php?title=Pipeline_(software)&oldid=660291081`

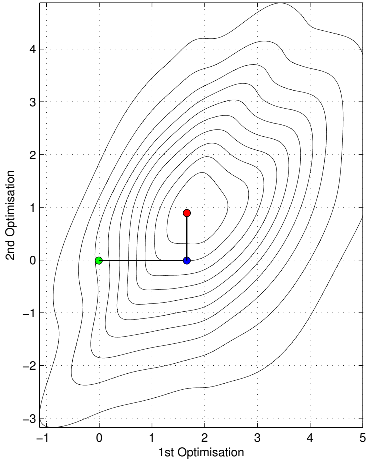
Optimising two parameters



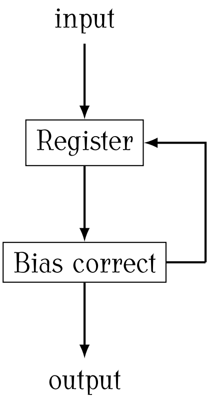


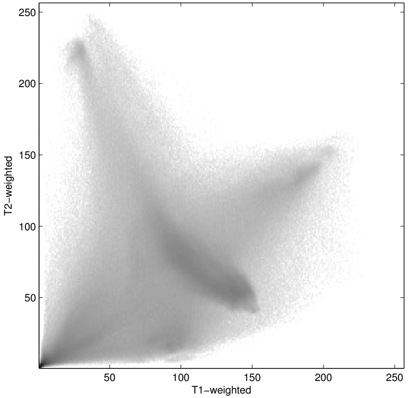
Single pass


2nd Optimisation
1st Optimisation
4
3
2
1
0
−1
−2
−3
−1
0
1
2
3
4
5


2nd Optimisation
1st Optimisation
4
3
2
1
0
−1
−2
−3
−1
0
1
2
3
4
5

Optimising two functions





Bottom-up & top-down processing in the brain

- Pipelines are a purely bottom up approach, with no top-down control.
- Data processing in the brain involves both top-down and bottom-up processing.
- Can not expect to achieve optimal understanding from a purely bottom-up approach.

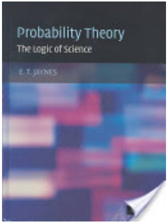
Generative Models

A generative model is a model for randomly generating observable data, typically given some hidden parameters. It specifies a joint probability distribution over observation and label sequences. Generative models are used in machine learning for either modeling data directly (i.e., modeling observations draws from a probability density function), or as an intermediate step to forming a conditional probability density function. A conditional distribution can be formed from a generative model through Bayes' rule.

Generative model. (2015, April 30). In Wikipedia, The Free Encyclopedia. Retrieved 16:46, June 17, 2015, from `https://en.wikipedia.org/w/index.php?title=Generative_model&oldid=660109222`

Probaility Theory

> *"Probability theory is nothing but common sense reduced to calculation."*
>
> — Laplace



Desiderata of probability theory:

1. Representation of degree of plausibility by real numbers.
2. Qualitative correspondence with common sense.
3. Consistency.

Jaynes, Edwin T. *Probability theory: the logic of science.* Cambridge university press, 2003.

Product and Sum Rules

Product Rule

$$\begin{aligned} p(\mathbf{y}, \mathbf{x}) &= p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \\ &= p(\mathbf{x}|\mathbf{y})p(\mathbf{y}) \end{aligned}$$

Sum Rule

$$p(\mathbf{y}) = \sum_{\mathbf{x}} p(\mathbf{y}, \mathbf{x})$$

or for continuous $\mathbf{x}$

$$p(\mathbf{y}) = \int_{\mathbf{x}} p(\mathbf{y}, \mathbf{x}) d\mathbf{x}$$

$p(\mathbf{x})$ is the probability of $\mathbf{x}$.
$p(\mathbf{x}, \mathbf{y})$ is the joint probability of $\mathbf{x}$ and $\mathbf{y}$.
$p(\mathbf{x}|\mathbf{y})$ is the probability of $\mathbf{x}$ conditional on $\mathbf{y}$.

Bayes Rule

Combining the sum and product rules, gives Bayes rule:

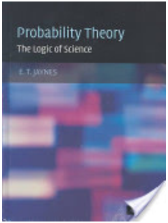
$$p(\theta|\mathbf{X}) = \frac{p(\mathbf{X}|\theta)p(\theta)}{p(\mathbf{X})} = \frac{p(\mathbf{X}|\theta)p(\theta)}{\int_\theta p(\mathbf{X}|\theta)p(\theta)d\theta}$$

In words:

$$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}}$$

Ignorance priors

- Sometimes we don't have previous observations to formulate priors.
- Jaynes suggests using a maximum entropy prior.
- An ignorance prior is a prior probability distribution where equal probability is assigned to all possibilities.
- Ignorance priors can be motivated via invariance/symmetry (transformation groups).

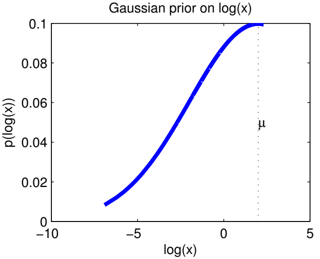


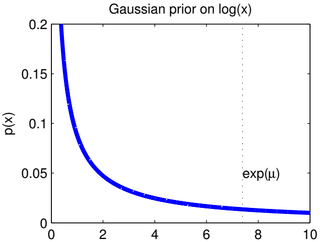
Jaynes, Edwin T. *Probability theory: the logic of science*. Cambridge university press, 2003.

Priors for positive values

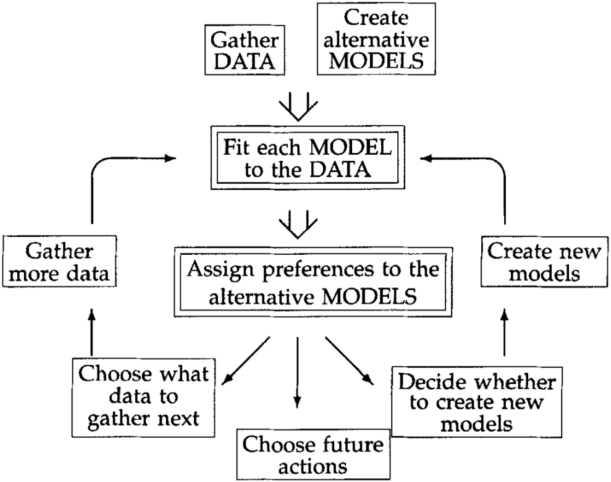
- Some things can not be less than zero.
    - Counts of observed photons.
    - Multiplicative “bias” fields.
    - Lengths, areas, volumes, etc.
- Formulate the model via logarithms, and impose a prior on these.

Jeffreys, Harold. *“An invariant form for the prior probability in estimation problems.”* In Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, vol. 186, no. 1007, pp. 453-461. The Royal Society, 1946.



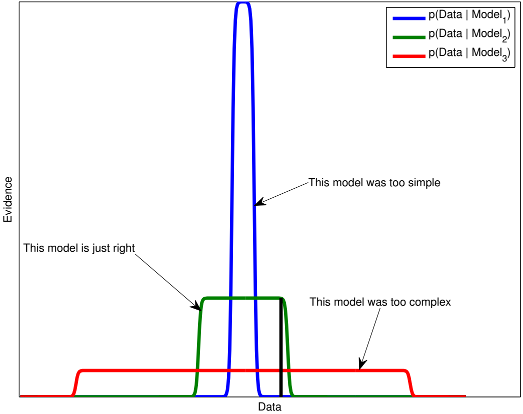


Scientific process



MacKay, David JC. "*Bayesian interpolation.*" Neural computation 4, no. 3 (1992): 415-447.

Goldilocks and the three Bayesian models



"Everything should be made as simple as possible, but not simpler."

— Einstein (possibly)

General aim of medical image computing

Given an image or a set of images $\mathbf{x}^*$, best predict $\mathbf{y}^*$.
Here, $\mathbf{y}$ may be:

- A diagnosis.
- An optimal treatment decision.
- Another image, for example:
  - A cleaned up version of the same image.
  - A map of where a neurosurgeon should best avoid.
  - A map of gamma ray absorption for attenuation correction in MR/PET.
- etc

General aim of medical image computing

Often a collection of training data to work from ($\mathbf{X}$ and $\mathbf{Y}$).
The aim becomes of of determining $p(\mathbf{y}^*|\mathbf{x}^*, \mathbf{Y}, \mathbf{X})$.

General aim of medical image computing

Predictions are based on some model, $\mathscr{M}$. Usually, a model has parameters, θ:

$$\begin{aligned} p(\mathbf{y}^*|\mathbf{x}^*, \mathbf{Y}, \mathbf{X}, \mathscr{M}) &= \int_\theta p(\mathbf{y}^*, \theta|\mathbf{x}^*, \mathbf{Y}, \mathbf{X}, \mathscr{M}) d\theta \\ &= \int_\theta p(\mathbf{y}^*|\mathbf{x}^*, \mathbf{Y}, \mathbf{X}, \theta, \mathscr{M}) p(\theta|\mathscr{M}) d\theta \end{aligned}$$

Predictions may also be made by averaging over models.

$$p(\mathbf{y}^*|\mathbf{x}^*, \mathbf{Y}, \mathbf{X}) = \sum_i p(\mathbf{y}^*|\mathbf{x}^*, \mathbf{Y}, \mathbf{X}, \mathscr{M}_i) P(\mathscr{M}_i)$$

Unfortunately...

"In theory, there is no difference between theory and practice. But, in practice, there is."

Many of the integrations needed to compute model evidence are not computationally feasible in medical image computing applications. Workarounds include:

- Use *maximum a posteriori* (MAP) estimation, and approximate probability distributions via a delta function.

$$\hat{\theta} = \arg\max_{\theta} \log p(\mathbf{X}, \theta)$$

- Model selection via cross-validation.

1 Introduction

2 Segmentation

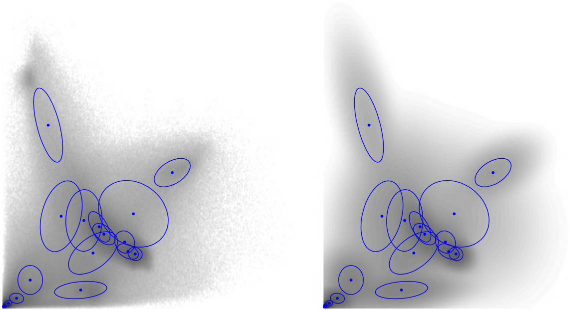
3 Diffeomorphic Registration

4 Longitudinal Registration

5 Dimensionality Reduction

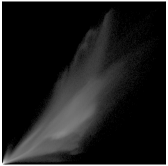
Mixture of Gausians

$$
\begin{aligned}
\mathcal{E} &= -\log p(\mathbf{f}|\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\gamma}) \\
&= -\sum_{i=1}^{I} \log \left(\sum_{k=1}^{K} \frac{\gamma_k}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{(f_i-\mu_k)^2}{2\sigma_k^2} \right) \right)
\end{aligned}
$$

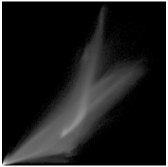


Incorporating "bias" correction

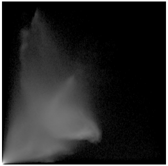
$$\mathcal{E} = -\sum_{i=1}^{I} \log \left(\sum_{k=1}^{K} \frac{\gamma_k}{\sqrt{2\pi \frac{\sigma_k^2}{\rho_i(\boldsymbol{\beta})^2}}} \exp\left(-\frac{\left(f_i - \frac{\mu_k}{\rho_i(\boldsymbol{\beta})}\right)^2}{2\frac{\sigma_k^2}{\rho_i(\boldsymbol{\beta})^2}} \right) \right)$$



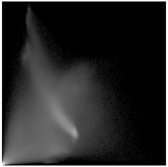
Original



Corrected



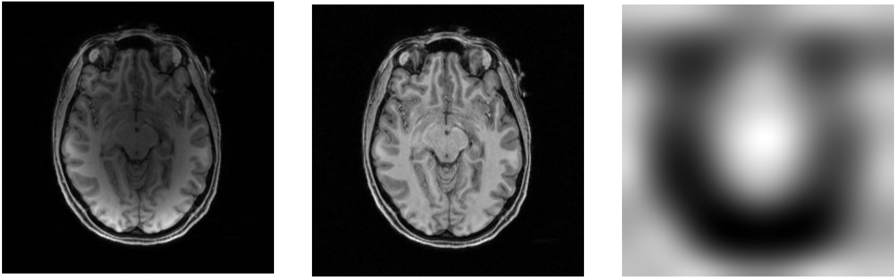
Original



Corrected

Incorporating "bias" correction

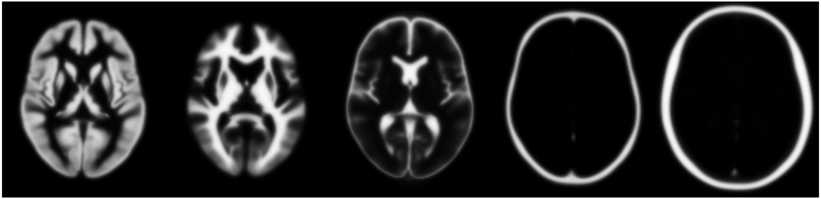
$$\mathcal{E} = -\sum_{i=1}^{I} \log \left(\rho_i(\boldsymbol{\beta}) \sum_{k=1}^{K} \frac{\gamma_k}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{(\rho_i(\boldsymbol{\beta}) f_i - \mu_k)^2}{2\sigma_k^2} \right) \right)$$



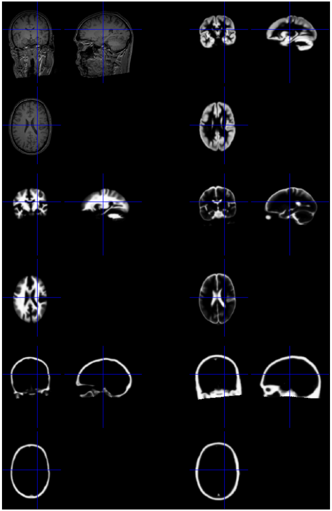
Original Corrected Field

Incorporating deformable tissue priors

$$\mathcal{E} = -\sum_{i=1}^{I} \log \left(\frac{\rho_i(\boldsymbol{\beta})}{\sum_{k=1}^{K} \gamma_k b_{ik}(\boldsymbol{\alpha})} \sum_{k=1}^{K} \frac{\gamma_k b_{ik}(\boldsymbol{\alpha})}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{(\rho_i(\boldsymbol{\beta}) f_i - \mu_k)^2}{2\sigma_k^2} \right) \right)$$

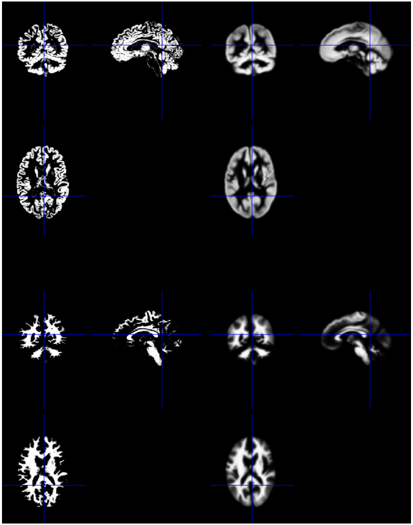


Incorporating deformable tissue priors



$$\mathcal{E} = -\sum_{i=1}^{I} \log \left(\frac{\rho_i(\beta)}{\sum_{k=1}^{K} \gamma_k b_{ik}(\alpha)} \sum_{k=1}^{K} \frac{\gamma_k b_{ik}(\alpha)}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{(\rho_i(\beta) f_i - \mu_k)^2}{2\sigma_k^2} \right) \right)$$

Latent variables



Optimisation done via EM.

Marginalised with respect to latent variables ($\mathbf{z}$), which encode tissue class memberships.

$$p(\mathbf{f}, \theta) = \int_{\mathbf{z}} p(\mathbf{f}, \mathbf{z}, \theta) dz$$

where

$$\theta = \{\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\alpha}\}$$

- Ashburner, John, and Karl J. Friston. "*Unified segmentation.*" Neuroimage 26, no. 3 (2005): 839-851.
- `http://www.fil.ion.ucl.ac.uk/spm/software/spm12/`, `spm12/spm_preproc_run.m`.

1 Introduction

2 Segmentation

3 Diffeomorphic Registration
- Groupwise registration
- LDDMM
- Shooting

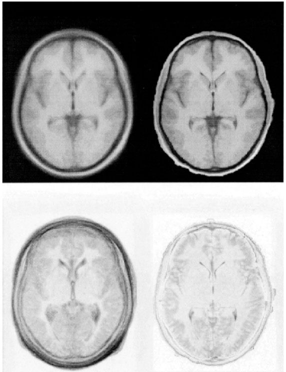
4 Longitudinal Registration

5 Dimensionality Reduction

"Groupwise registration"

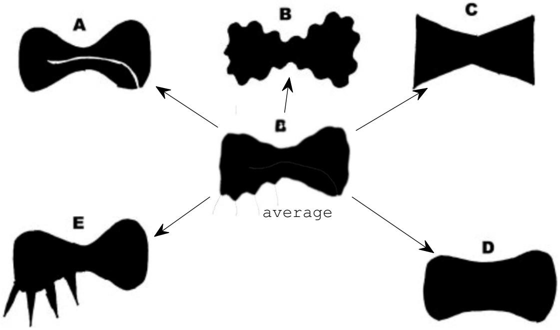
Sometimes the aim is to align multiple scans together.
Ignoring the many technical details, the procedure involves alternating between:

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



An early attempt (1999).

"Groupwise registration"



"Groupwise registration"

"Groupwise registration"

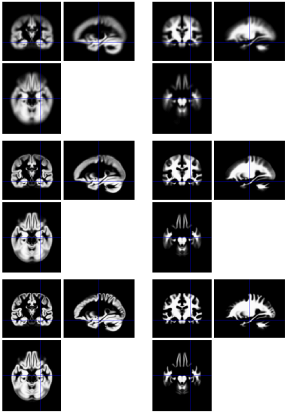
Based on matching K tissue maps together via a multinomial model, where:

$$\log P(\mathbf{f}(\mathbf{x})|\boldsymbol{\mu},\varphi) = \sum_{k=1}^{K} f_k(\mathbf{x}) \log \mu_k(\varphi(\mathbf{x}))$$

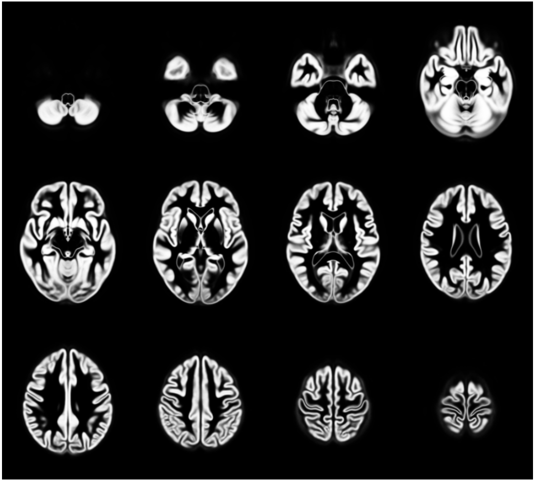
Tissue probabilities sum to 1 at each voxel:

$$\mu_k \geq 0, \sum_{k=1}^{K} \mu_k = 1$$

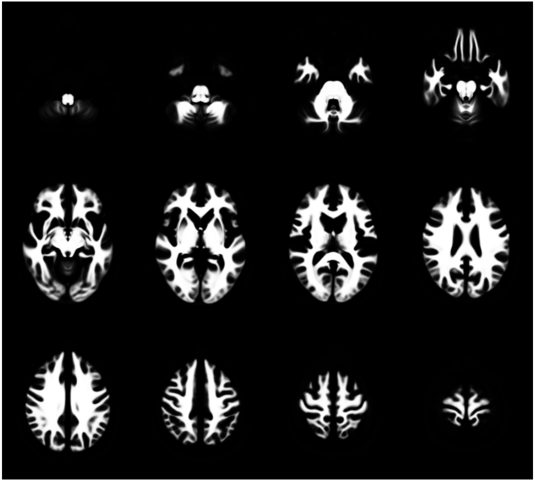
$$f_k \geq 0, \sum_{k=1}^{K} f_k = 1$$



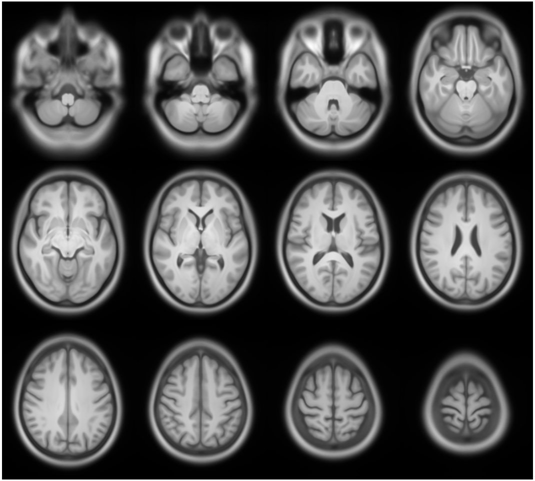
“Groupwise registration”



"Groupwise registration"



"Groupwise registration"



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

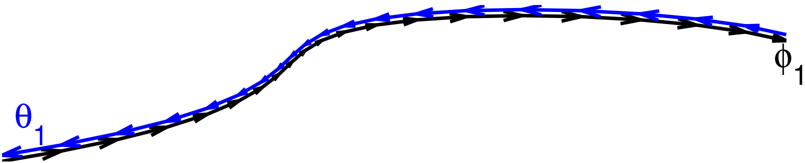
Metric distances

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).

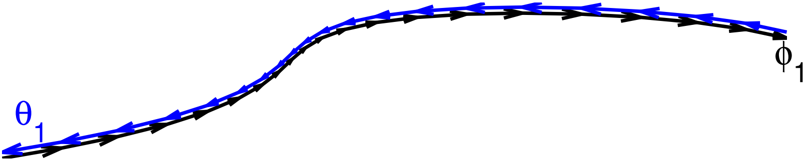
Satisfying (3) requires inverse-consistent image registration.
Satisfying (4) requires a specific class of image registration models.

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:

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

For large N, the evolution of a deformation may be conceptualised as integrating the following equation:

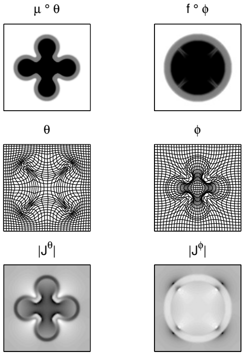
$$\frac{d\boldsymbol{\varphi}}{dt} = \mathbf{v}_t(\boldsymbol{\varphi})$$

Geodesic distances (from zero) are then measured by:

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

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:

$$\mathcal{E} = \frac{1}{2}\int_{t=0}^{1} ||\mathbf{L}\mathbf{v}_t||^2 dt + \frac{1}{2\sigma^2}||f - \mu\left(\boldsymbol{\varphi}_1^{-1}\right)||^2$$

$$\text{where } \boldsymbol{\varphi}_0 = \text{id}, \frac{d\boldsymbol{\varphi}}{dt} = \mathbf{v}_t\left(\boldsymbol{\varphi}_t\right)$$

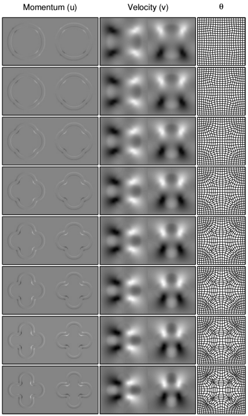
First term is a squared deformation distance measure.
Second term is the squared difference between images.
The objective is to estimate a series of velocity fields $(\mathbf{v}_t)$.

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).

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 ($\boldsymbol{\varphi}_0 = id$), and then evolve the following dynamical system for unit time:

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

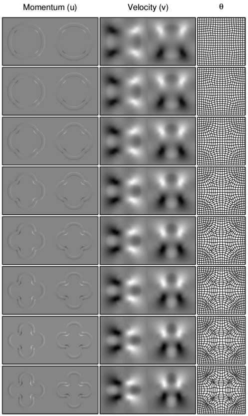


Younes, L, Arrate, F & Miller, MI. *Evolutions equations in computational anatomy.* Neuroimage 45(1S1):40–50 (2009).

LDDMM via "geodesic shooting"

The final deformation ($\boldsymbol{\varphi}_1$) 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`



Younes, L, Arrate, F & Miller, MI. *Evolutions equations in computational anatomy.* Neuroimage 45(1S1):40–50 (2009).

"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:

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

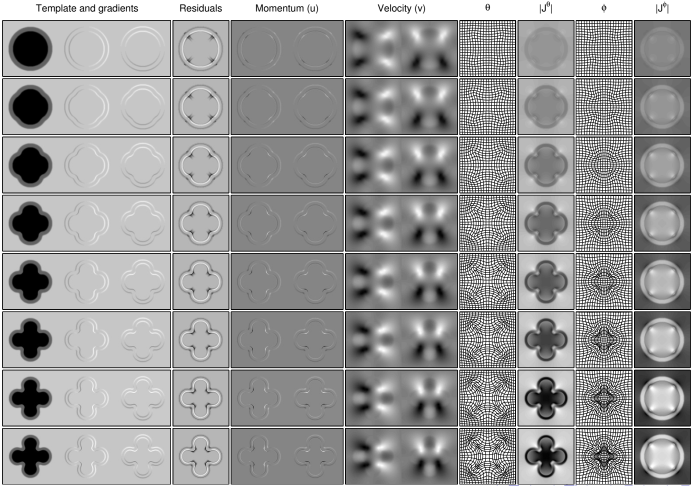
"Scalar Momentum"

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

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"*, $\det|\mathbf{D}\boldsymbol{\varphi}_1|(\mu - f \circ \boldsymbol{\varphi}_1)$. This is a scalar field, and in principle is all that is needed (along with the template) to reconstruct the original images.

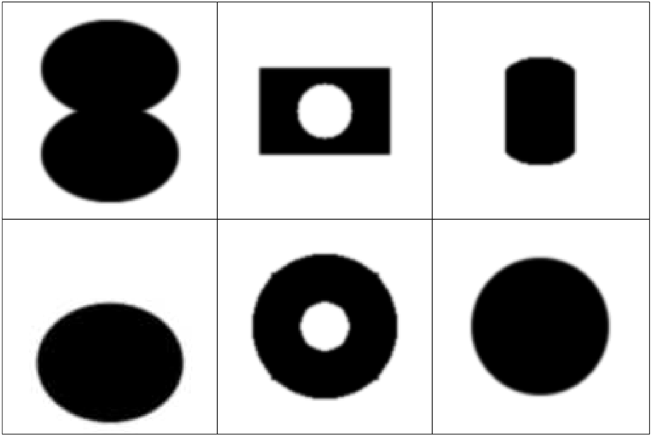
Miller et al. "Collaborative computational anatomy: an MRI morphometry study of the human brain via diffeomorphic metric mapping." Human Brain Mapping 30(7):2132–2141 (2009).
Singh, Fletcher, Preston, Ha, King, Marron, Wiener & Joshi (2010). *Multivariate Statistical Analysis of Deformation Momenta Relating 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|

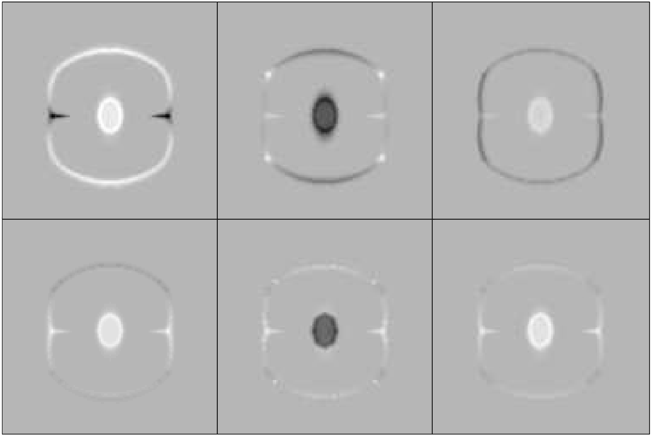
Example Images

Some example images.



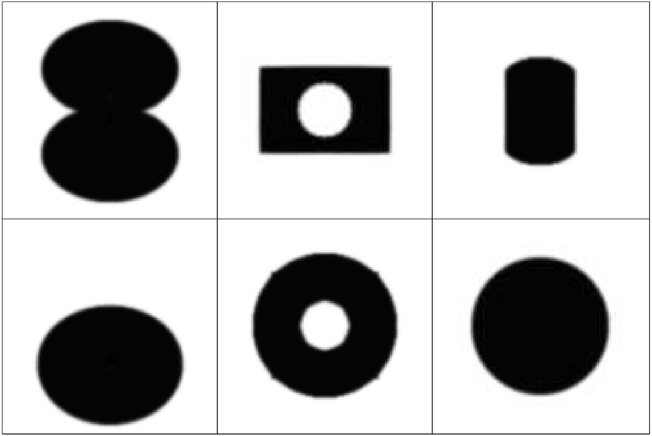
Scalar Momentum

Scalar momenta after aligning to a common template.



Reconstructed Images

Images reconstructed from scalar momenta (and template).



"Shapes are the ultimate non-linear sort of thing"

David Mumford

- Ashburner, John, and Karl J. Friston. *"Diffeomorphic registration using geodesic shooting and Gauss-Newton optimisation."* NeuroImage 55, no. 3 (2011): 954-967.
- Ashburner, John, and Stefan Klöppel. *"Multivariate models of inter-subject anatomical variability."* Neuroimage 56, no. 2 (2011): 422-439.
- Ashburner, John, and Michael I Miller. *"Diffeomorphic Image Registration."* In *Brain Mapping: an Encyclopedic Reference*, pp. 315-321. Academic Press: Elsevier (2015). Toga AW (ed.).
- `http://www.fil.ion.ucl.ac.uk/spm/software/spm12/`, `spm12/toolbox/Shoot`.

1 Introduction

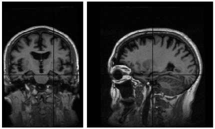
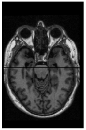
2 Segmentation

3 Diffeomorphic Registration

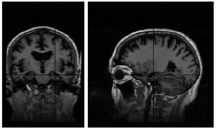
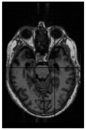
4 Longitudinal Registration
- Bias Fields
- Rigid-Body
- Diffeomorphisms
- Combined Model

5 Dimensionality Reduction

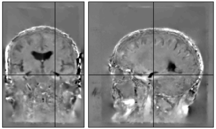
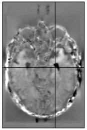
Longitudinal data: OAS2_0002

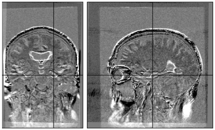
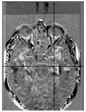
First scan (0 days) after rigid body registration.

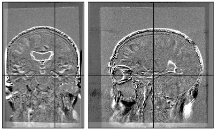
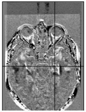
Second scan (1895 days) after rigid body registration.

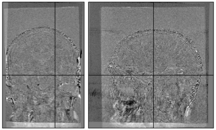
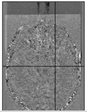
Jacobian determinants windowed between 0.7 and 1.3.

Residuals from rigid body registration, without bias correction.

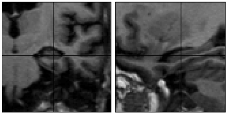
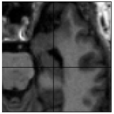



Residuals from rigid body registration, with bias correction.

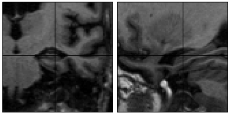
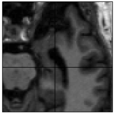



Residuals from diffeomorphic registration, with bias correction.

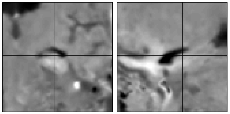
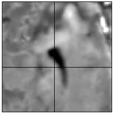
Longitudinal data: OAS2_0002

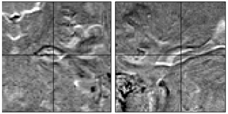
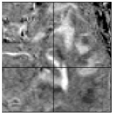
First scan (0 days)
after rigid body
registration.

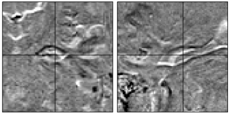
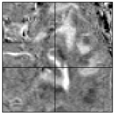
Second scan (1895 days)
after rigid body
registration.

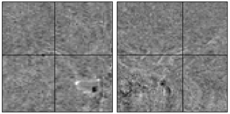
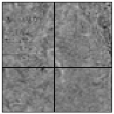
Jacobian determinants
windowed between
0.7 and 1.3.

Residuals from
rigid body registration,
without bias correction.

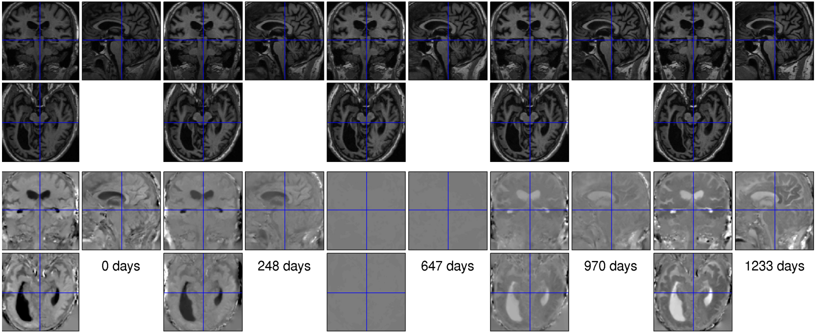



Residuals from
rigid body registration,
with bias correction.

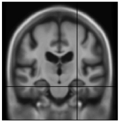
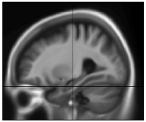
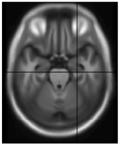



Residuals from
diffeomorphic
registration,
with bias correction.

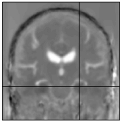
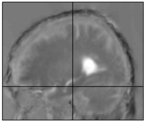
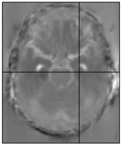
Longitudinal data: OAS2_0048



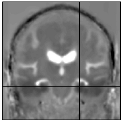
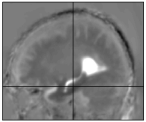
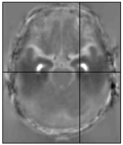
Longitudinal data: Averages

Mean image.

Control subjects expansion/year.

Dementia subjects expansion/year.

Optimisation

Problem is treated as finding a maximum a posteriori (or regularised maximum likelihood) solution.

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \mathcal{E}(\boldsymbol{\theta})$$

where

$$\mathcal{E}(\boldsymbol{\theta}) \equiv -\log p(\boldsymbol{\theta}, \mathrm{Data}) = -\log p(\mathrm{Data}|\boldsymbol{\theta}) - \log p(\boldsymbol{\theta})$$

Optimisation: Newton's Method

An iterative local optimisation scheme:

$$\begin{aligned}
\boldsymbol{\theta}^{(n+1)} &= \boldsymbol{\theta}^{(n)} - \left[\mathbf{H}\left(\mathcal{E}(\boldsymbol{\theta}^{(n)})\right)\right]^{-1} \nabla\mathcal{E}(\boldsymbol{\theta}^{(n)}) \\
\text{where } \mathbf{H}\left(\mathcal{E}(\boldsymbol{\theta}^{(n)})\right) &= \text{Hessian matrix of 2nd derivatives} \\
\nabla\mathcal{E}(\boldsymbol{\theta}^{(n)}) &= \text{vector of 1st derivatives}
\end{aligned}$$

Note: may converge to a maximum, minimum or saddle point, depending on whether or not the Hessian is positive definite.

Optimisation: Gauss-Newton Algorithm

Gauss-Newton method can be used for least-squares minimisation, where the objective function has the following form:

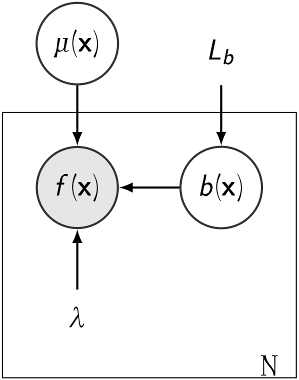
$$\mathcal{E}(\boldsymbol{\theta}) = \sum_i r_i^2(\boldsymbol{\theta})$$

- Ensures a positive definite approximation of the Hessian.
- Converges (hopefully) to a local minimum.

$$\begin{aligned}\boldsymbol{\theta}^{(n+1)} &= \boldsymbol{\theta}^{(n)} - \left[\mathbf{J}^T\mathbf{J}\right]^{-1}\mathbf{J}^T\mathbf{r} \\ \text{where } \mathbf{J} &= \frac{\partial r_i}{\partial \theta_j}(\boldsymbol{\theta}^{(n)})\end{aligned}$$

Can also motivate a positive definite approximation via a Fisher information matrix (as in Fisher scoring).

Bias Fields: Generative Model



$f(\mathbf{x})$ – image

$\mu(\mathbf{x})$ – mean image

λ – noise precision

$b(\mathbf{x})$ – bias field

L_b – bias regularisation

N – number of images

Bias Fields: Objective Function

Minimise the following:

$$\mathcal{E} = \sum_{n=1}^{N} \left(\frac{\lambda_n}{2} \| f_n - \mu e^{b_n} \|^2 + \frac{1}{2} \| L_b b_n \|^2 \right)$$

f – image

μ – mean image

λ – noise precision

b – bias field

L_b – bias regularisation

N – number of images

Bias Fields: Exponential Map

- The "bias field" is really not a bias, as it is multiplicative rather than additive.
- Want the probability of re-scaling by (say) 2 to be the same as that of scaling by $\frac{1}{2}$.
- Parameterise by a field $b(\mathbf{x})$, and generate bias from the exponential.

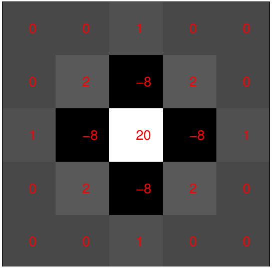
$$\exp(b) = \lim_{n\to\infty}\left(1 + \frac{b}{n}\right)^n$$

Bias Fields: Regularisation

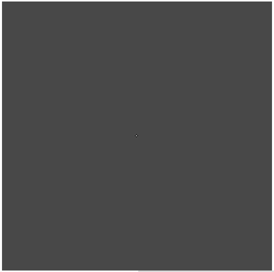
Penalise sum of squares of second derivatives:

$$\|L_b b\|^2 = \omega_0 \int_{\mathbf{x}} \|\nabla^2 b(\mathbf{x})\|^2 d\mathbf{x}$$

Differential operator ($L_b^{\dagger} L_b$)

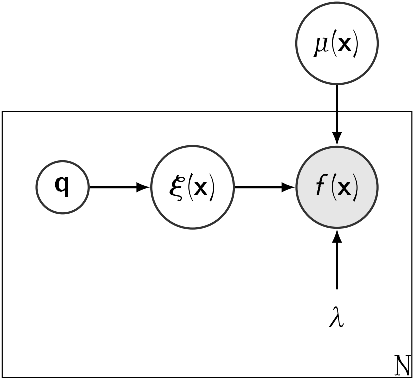


Differential operator (zoomed out)



Green's function (via FFTs)

Rigid-Body: Generative Model



$f(\mathbf{x})$ – image

$\mu(\mathbf{x})$ – mean image

λ – noise precision

$\xi(\mathbf{x})$ – rigid-body transform

$\mathbf{q}$ – rigid-body parameters

N – number of images

Rigid-Body: Objective Function

$$\mathcal{E} = \sum_{n=1}^{N} \frac{\lambda_n}{2} \| f_n - \mu(\xi_{\mathbf{q}_n}^{-1}) \|^2 = \sum_{n=1}^{N} \frac{\lambda_n}{2} \int_{\mathbf{x}} |\mathbf{D}\xi_{\mathbf{q}_n}(\mathbf{x})| (f_n(\xi_{\mathbf{q}_n}(\mathbf{x})) - \mu(\mathbf{x}))^2 d\mathbf{x}$$

- f – image
- μ – mean image
- λ – noise precision
- $\xi_{\mathbf{q}}$ – rigid-body transform
- N – number of images

Note the change of variables.

$$\int_{\mathbf{x}} g(\mathbf{x}) d\mathbf{x} = \int_{\mathbf{x}} g(\varphi(\mathbf{x})) |\mathbf{D}\varphi(\mathbf{x})| d\mathbf{x}$$

where $|\mathbf{D}\varphi(\mathbf{x})|$ means the Jacobian determinant of φ at $\mathbf{x}$.

Rigid-Body: Exponential Map

A rigid-body transformation matrix ($\mathbf{R}_\mathbf{q} \in SE(3)$) is computed via a matrix exponential:

$$\mathbf{R}_\mathbf{q} = \exp \begin{bmatrix} 0 & q_4 & -q_5 & q_1 \\ -q_4 & 0 & q_6 & q_2 \\ q_5 & -q_6 & 0 & q_3 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \text{ where } \exp \mathbf{Q} = \sum_{n=0}^{\infty} \frac{1}{n!} \mathbf{Q}^n.$$

A mapping from each voxel in the template, to the corresponing voxel in the nth image is by:

$$\xi_{\mathbf{q}_n}(\mathbf{x}) = \mathbf{I}_{3,4} \mathbf{M}_n^{-1} \mathbf{R}_{\mathbf{q}_n} \mathbf{M}_\mu \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix}, \text{ where } \mathbf{I}_{3,4} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Each $\mathbf{M}$ maps from voxels to corresponding mm coordinates.

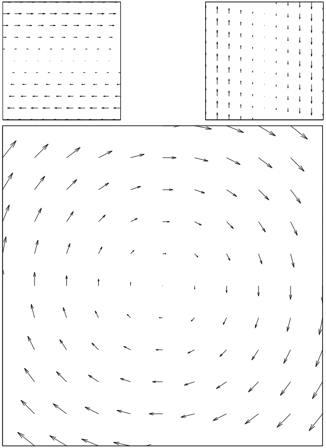
Rigid-Body: Exponential Map

Rotation in 2D ($\mathbf{R_q} \in SO(2)$):

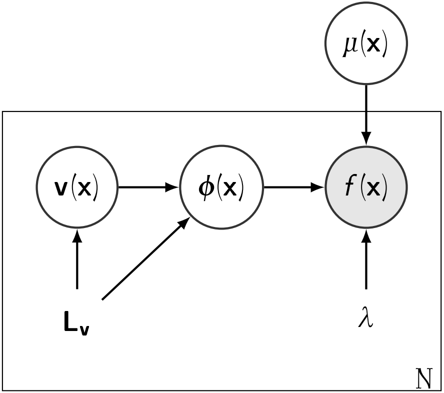
$$\mathbf{R_q} = \exp \begin{bmatrix} 0 & q_1 \\ -q_1 & 0 \end{bmatrix}$$

Computing a matrix exponential is analagous to integrating a dynamical system over unit time.

$$\mathbf{R_q} = \lim_{n\to\infty} \begin{bmatrix} 1 & q_1/n \\ -q_1/n & 1 \end{bmatrix}^n$$



Diffeomorphisms: Generative Model



$f(\mathbf{x})$ – image

$\mu(\mathbf{x})$ – mean image

λ – noise precision

$\phi(\mathbf{x})$ – diffeomorphism

$\mathbf{v}(\mathbf{x})$ – initial velocity

$\mathbf{L_v}$ – velocity regularisation

N – number of images

Diffeomorphisms: Objective Function

$$\mathcal{E} = \sum_{n=1}^{N} \left(\frac{\lambda_n}{2} \| f_n - \mu \circ \phi_{\mathbf{v}_n}^{-1} \|^2 + \frac{1}{2} \| \mathbf{L}_{\mathbf{v}_n} \mathbf{v}_n \|^2 \right)$$

$$= \sum_{n=1}^{N} \left(\frac{\lambda_n}{2} \int_{\mathbf{x}} |\mathbf{D}\phi_{\mathbf{v}_n}(\mathbf{x})| (f_n(\phi_{\mathbf{v}_n}(\mathbf{x})) - \mu(\mathbf{x}))^2 d\mathbf{x} + \frac{1}{2} \| \mathbf{L}_{\mathbf{v}_n} \mathbf{v}_n \|^2 \right)$$

f – image

μ – mean image

λ – noise precision

$\phi_{\mathbf{v}}$ – diffeomorphism

$\mathbf{v}$ – velocity field

$\mathbf{L}_{\mathbf{v}}$ – velocity regularisation

Note: Diffeomorphic deformations are computed via a Riemannian exponential.

Diffeomorphisms: Exponential Map

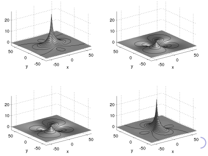
Riemannian exponantial is computed via geodesic shooting. Initialise $\phi_{\mathbf{v}}$ to the identity transform and compute initial momentum from initial velocity via:

$$\mathbf{u} = \mathbf{L}_{\mathbf{v}}^{\dagger}\mathbf{L}_{\mathbf{v}}\mathbf{v}.$$

Then the following dynamical system is integrated over unit time:

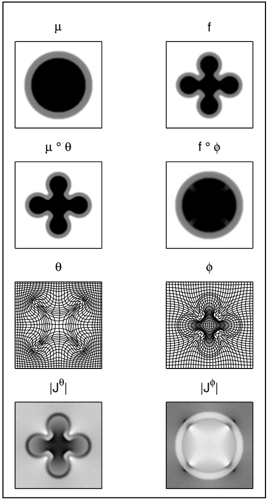
$$\dot{\phi}_{\mathbf{v}} = \left(\mathbf{K}_{\mathbf{v}}\left(\left|\mathbf{D}\phi_{\mathbf{v}}^{-1}\right|\left(\mathbf{D}\phi_{\mathbf{v}}^{-1}\right)^{T}\left(\mathbf{u}\circ\phi_{\mathbf{v}}^{-1}\right)\right)\right)\circ\phi_{\mathbf{v}}$$

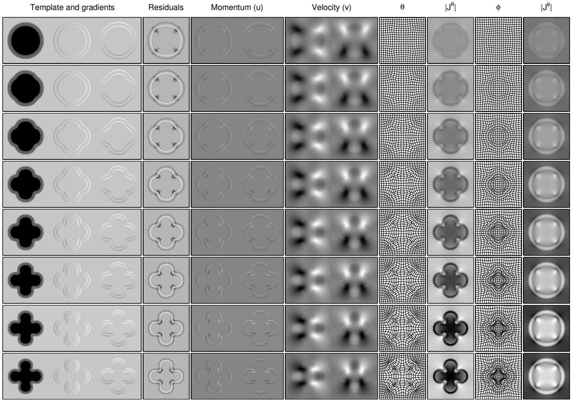
$\mathbf{K}_{\mathbf{v}}$ is the Green's function of $\mathbf{L}_{\mathbf{v}}^{\dagger}\mathbf{L}_{\mathbf{v}}$, such that:



$$\begin{aligned}\mathbf{K}_{\mathbf{v}}\mathbf{L}_{\mathbf{v}}^{\dagger}\mathbf{L}_{\mathbf{v}}\mathbf{v} &= \mathbf{v}\\ \mathbf{L}_{\mathbf{v}}^{\dagger}\mathbf{L}_{\mathbf{v}}\mathbf{v}\mathbf{K}_{\mathbf{v}}\mathbf{u} &= \mathbf{u}\end{aligned}$$

Diffeomorphisms: Exponential Map





$$\dot{\phi}_{\mathbf{v}} = \left(\mathbf{K}_{\mathbf{v}}\left(\left|\mathbf{D}\phi_{\mathbf{v}}^{-1}\right|\left(\mathbf{D}\phi_{\mathbf{v}}^{-1}\right)^{T}\left(\mathbf{u}\circ\phi_{\mathbf{v}}^{-1}\right)\right)\right)\circ\phi_{\mathbf{v}}$$

Diffeomorphisms: Regularisation

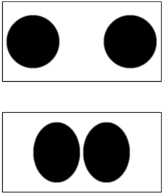
$$\|\mathbf{L}_\mathbf{v}\mathbf{v}\|^2 = \int_\mathbf{x} \left(\frac{\omega_1}{4} \|\mathbf{Dv} + (\mathbf{Dv})^T\|_F^2 + \omega_2 \mathrm{tr}(\mathbf{Dv})^2 + \omega_3 \|\nabla^2 \mathbf{v}\|^2 \right) d\mathbf{x}$$

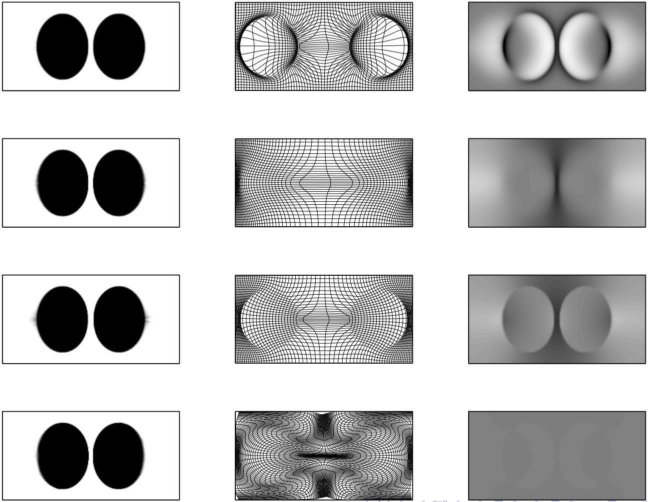
Three hyper-parameters are involved:

- ω_1 controls stretching and shearing (but not rotation).
- ω_2 controls the divergence, which in turn determines the amount of volumetric expansion and contraction.
- ω_3 controls the bending energy. This ensures that the resulting velocity fields have smooth spatial derivatives.

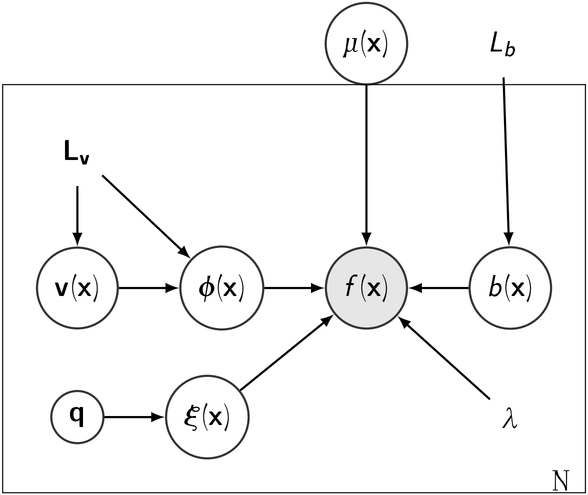
Diffeomorphisms: Regularisation

Two simulated images





Combined Model: Generative Model



- $\boldsymbol{f}(\mathbf{x})$ – image
- μ – mean image
- λ – noise precision
- $\boldsymbol{b}(\mathbf{x})$ – "bias" field
- $\boldsymbol{L_b}$ – bias field regularisation
- $\xi(\mathbf{x})$ – rigid-body transform
- $\mathbf{q}$ – rigid-body parameters
- $\phi_{\mathbf{v}}$ – diffeomorphism
- $\mathbf{v}$ – velocity field
- $\mathbf{L_v}$ – velocity regularisation
- N – number of images

Combined Model: Generative Model

Minimise the following objective function:

$$\mathcal{E} = \sum_{n=1}^{N} \frac{1}{2} \int_{\mathbf{x}} \lambda_n \left|\mathbf{D}\boldsymbol{\varphi}_n(\mathbf{x})\right| \left(f_n'(\mathbf{x}) - \mu(\mathbf{x}) e^{b_n'(\mathbf{x})} \right)^2 dx + \sum_{n=1}^{N} \frac{1}{2} \left\| \mathbf{L}_{\mathbf{v}_n} \mathbf{v}_n \right\|^2 + \sum_{n=1}^{N} \frac{1}{2} \left\| L_b b_n \right\|^2$$

where:

$$\begin{aligned} \boldsymbol{\varphi}_n &= \xi_{\mathbf{q}_n} \circ \phi_{\mathbf{v}_n} \\ f_n' &= f_n(\boldsymbol{\varphi}_n) \\ b_n' &= b_n(\boldsymbol{\varphi}_n) \end{aligned}$$

"Everything is the way it is because it got that way"

D'Arcy Wentworth Thompson (1860–1948)

- Ashburner, John, and Gerard R. Ridgway. *"Symmetric diffeomorphic modeling of longitudinal structural MRI."* Frontiers in neuroscience 6 (2012).
- `http://www.fil.ion.ucl.ac.uk/spm/software/spm12/`, `spm12/toolbox/Longitudinal`.

1 Introduction

2 Segmentation

3 Diffeomorphic Registration

4 Longitudinal Registration

5 Dimensionality Reduction

Dimensionality reduction

- Kernel methods can be useful for relatively small datasets.
- Less useful for big big data.
    - $N \times N$ kernel matrix too large for memory.
    - May need to retain “horizontal” privacy in situations where patient data are mined across hospitals.
- Reduce dimensionality, while retaining as much information as possible.
- Construct some form of generative model.

Principal Component Analysis

Minimise the following w.r.t. $\mathbf{H}$ and $\mathbf{W}$:

$$\mathcal{E} = \sum_{n=1}^{N} \frac{1}{2}||\mathbf{f}_n - \sum_{k=1}^{K} \mathbf{h}_k w_{kn}||^2$$

Or this, w.r.t. $\boldsymbol{\mu}$, $\mathbf{H}$ and $\mathbf{W}$:

$$\mathcal{E} = \sum_{n=1}^{N} \frac{1}{2}||\mathbf{f}_n - \boldsymbol{\mu} - \sum_{k=1}^{K} \mathbf{h}_k w_{kn}||^2$$

EM for Principal Component Analysis

Given a $P \times N$ matrix $\mathbf{F}$, decompose it into a $P \times K$ matrix $\mathbf{H}$ and a $K \times N$ matrix $\mathbf{W}$, such that:

$$\mathbf{F} \simeq \mathbf{H}\mathbf{W}$$

The EM algorithm is:

- **E-step:** $\mathbf{W} \leftarrow (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{F}$
- **M-step:** $\mathbf{H} \leftarrow \mathbf{F}\mathbf{W}^T(\mathbf{W}\mathbf{W}^T)^{-1}$

Roweis, Sam. "*EM algorithms for PCA and SPCA.*" Advances in neural information processing systems (1998): 626-632.

Non-negative matrix factorisation

One form of NMF minimises the Frobenius Norm:

$$\mathcal{E} = \sum_{n=1}^{N} \frac{1}{2} ||\mathbf{f}_n - \sum_{k=1}^{K} \mathbf{h}_k w_{kn}||^2 \text{ , } \mathbf{W} \in \mathcal{R}_+^{K \times N} 0 \text{ , } \mathbf{H} \in \mathcal{R}_+^{P \times K}$$

The EM algorithm is similar, except it involves non-negative least squares (quadratic programming).

Lee, Daniel D., and H. Sebastian Seung. *"Algorithms for non-negative matrix factorization."* In Advances in neural information processing systems, pp. 556-562. 2001.
Lee, Daniel D., and H. Sebastian Seung. *"Learning the parts of objects by non-negative matrix factorization."* Nature 401, no. 6755 (1999): 788-791.

Generalised Principal Component Analysis

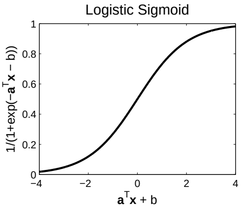
If $\mathbf{F}$ is binary, we could fit a logistic version by minimising the following w.r.t. $\mathbf{H}$ and $\mathbf{W}$:

$$\mathcal{E} = -\sum_{n=1}^{N}\sum_{p=1}^{P} \log(\sigma_{pn}) f_{pn} + \log(1-\sigma_{pn})(1-f_{pn})$$

where

$$\sigma_{pn} = \frac{1}{1+\exp(\sum_{k=1}^{K} h_{pk} w_{kn})}$$

The EM algorithm involves logistic regression.



Principal Geodesic Analysis

Could combine diffeomorphic registration with PCA by minimising:

$$\mathcal{E} = \sum_{n=1}^{N} \frac{\lambda}{2} ||\mathbf{f}_n - \boldsymbol{\mu} \circ \boldsymbol{\varphi}_n^{-1}||^2 + \frac{1}{2}||\mathbf{v_n}||_V^2$$

where $\mathbf{H}$ encodes principal components of initial velocity for computing diffeomorphisms:

$$\mathbf{v}_n = \sum_{k=1}^{K} \mathbf{h}_k w_{kn}$$

$$\boldsymbol{\varphi}_n = Exp(\mathbf{v}_n) \text{ (via geodesic shooting)}$$

Zhang, Miaomiao, and P. Thomas Fletcher. "*Probabilistic principal geodesic analysis.*" In Advances in Neural Information Processing Systems, pp. 1178-1186. 2013.
Zhang, Miaomiao, and P. Thomas Fletcher. "*Bayesian Principal Geodesic Analysis for Estimating Intrinsic Diffeomorphic Image Variability.*" Medical Image Analysis (2015).

Combined PCA/PGA model

Could combine diffeomorphic registration with PCA by minimising the following w.r.t. $\boldsymbol{\mu}$, $\mathbf{H}$, $\mathbf{A}$ and $\mathbf{W}$:

$$\mathcal{E} = \sum_{n=1}^{N} \frac{\lambda_1}{2} ||\mathbf{f}_n - (\boldsymbol{\mu} + \mathbf{r}_n) \circ \boldsymbol{\varphi}_n^{-1}||^2 + \frac{\lambda_2}{2} ||\mathbf{r}_n||^2 + \frac{1}{2} ||\mathbf{v}_n||_V^2$$

where:

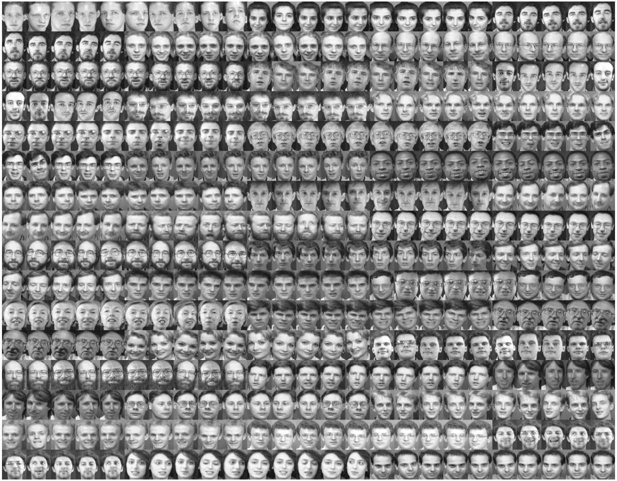
$$\mathbf{v}_n = \sum_{k=1}^{K} \mathbf{h}_k w_{kn}$$
$$\boldsymbol{\varphi}_n = Exp(\mathbf{v}_n)$$
$$\mathbf{r}_n = \sum_{k=1}^{K} \mathbf{a}_k w_{kn}$$

Note: Some form of *metamorphoses* approach may be better.

Richardson, Casey L., and Laurent Younes. "*Metamorphosis of Images in Reproducing Kernel Hilbert Spaces.*" arXiv preprint arXiv:1409.6573 (2014).

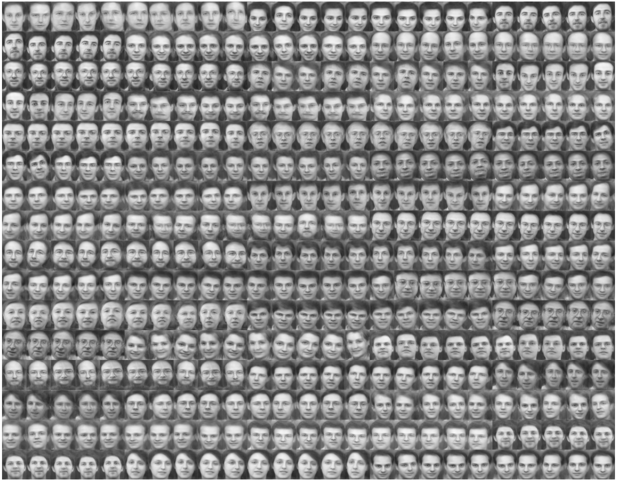
Original images

400 face images.



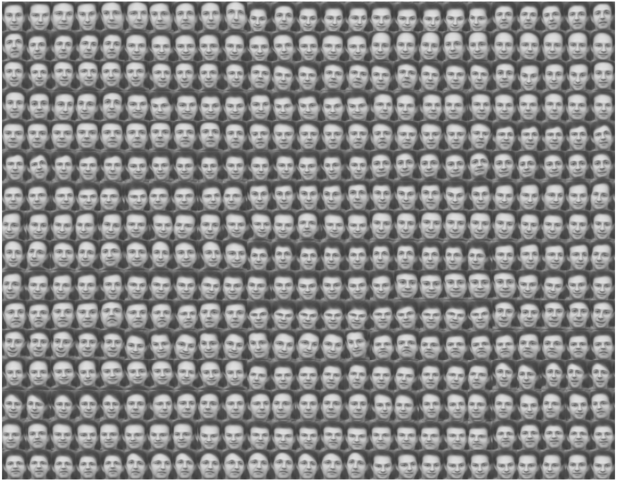
Shape and appearance model

Reconstruction with $K = 64$.



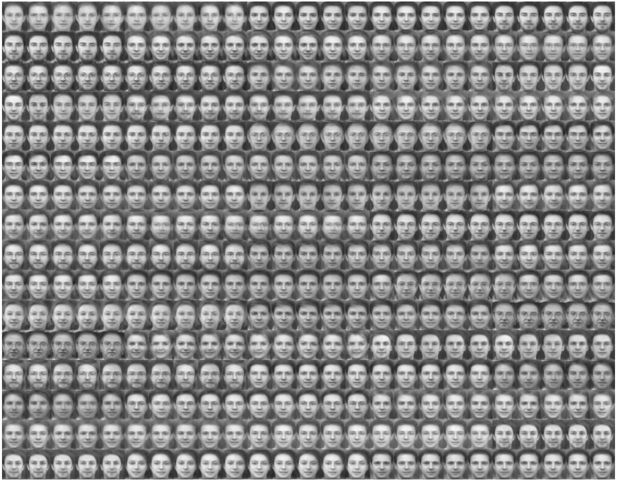
Shape model only

Ignoring the appearance variations.



Appearance model only

Ignoring the shape variations.

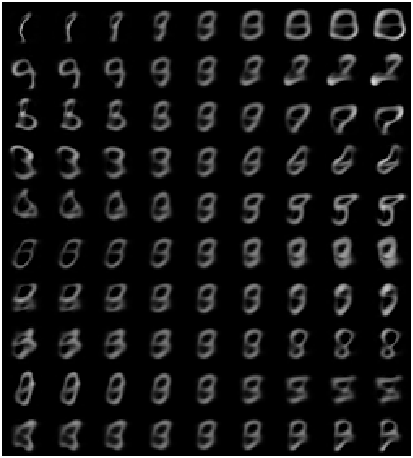


Eigen-components

Random samples

Random samples

MNIST Eigen-components

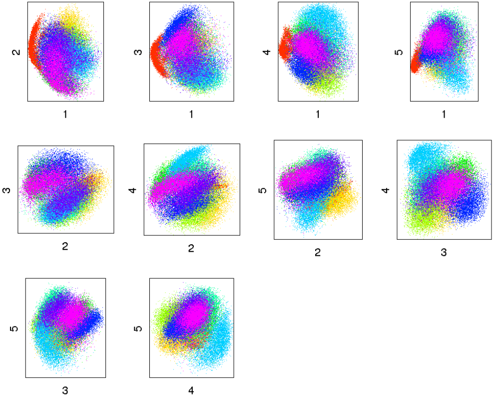


Yann LeCun, Corinna Cortes & Christopher J.C. Burges. http://yann.lecun.com/exdb/mnist/

MNIST Random samples

Yann LeCun, Corinna Cortes & Christopher J.C. Burges. http://yann.lecun.com/exdb/mnist/

MNIST Weights



Yann LeCun, Corinna Cortes & Christopher J.C. Burges. `http://yann.lecun.com/exdb/mnist/`

"To recognize shapes, first learn to generate images"

Geoffrey E Hinton (2007)


For the harmony of
the world is made
manifest in Form and
Number, and the heart
and soul and all the
poetry of Natural
Philosophy are embodied
in the concept of
mathematical beauty.
D'Arcy Thompson
On Growth and Form
(Dundee, 1917)