Generative Models for Medical Imaging

John Ashburner

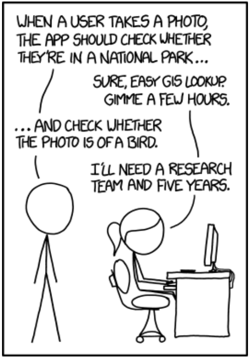
Wellcome Trust Centre for Neuroimaging,
UCL Institute of Neurology,
12 Queen Square,
London WC1N 3BG,
UK.

1. Introduction
   - Pipelines v Models
   - Probability Theory
   - Medical image computing
2. Segmentation
3. Longitudinal Registration

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



IN CS, IT CAN BE HARD TO EXPLAIN THE DIFFERENCE BETWEEN THE EASY AND THE VIRTUALLY IMPOSSIBLE.

https://xkcd.com

Why Image Processing Seems Easy

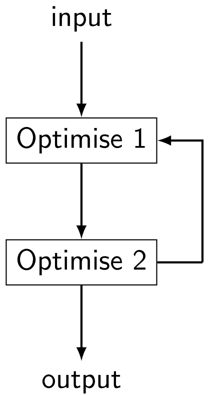
Neurons for visual processing take up 30% of the brain's cortex (as opposed to about 8% for touch and 3% for hearing).

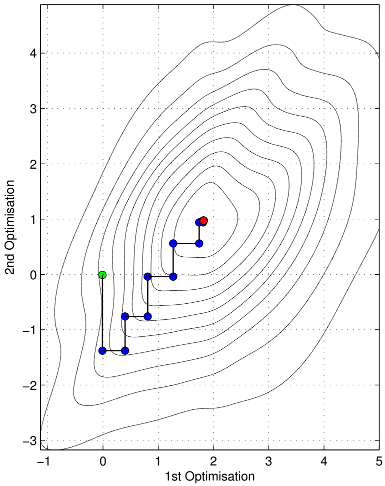
Pipelines

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

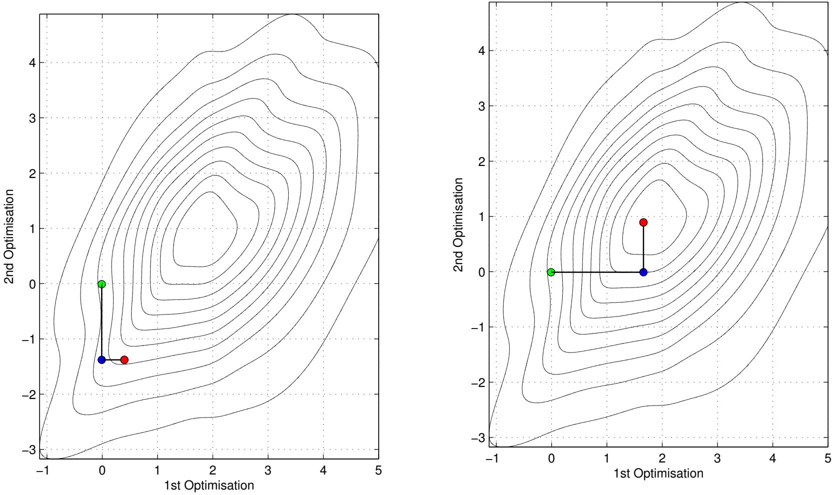
Pipeline (software). (2015, May 1). In Wikipedia, The Free Encyclopedia. Retrieved 16:50, June 17, 2015, from https://en.wikipedia.org/w/index.php?title=Pipeline_(software)&oldid=660291081

Optimising two parameters

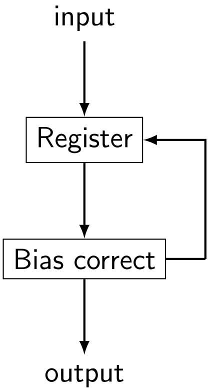


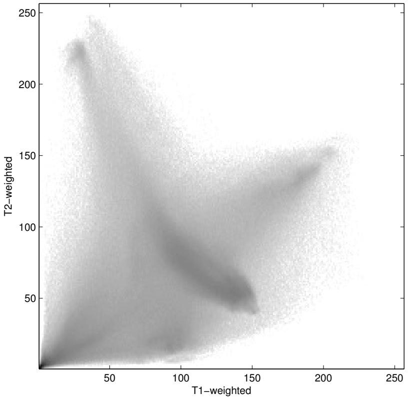


Single pass


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

Optimising two functions





Generative Models

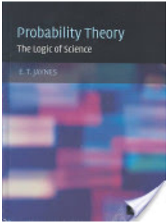
A generative model is a model for randomly generating observable data, typically given some hidden parameters. It specifies a joint probability distribution over observation and label sequences. Generative models are used in machine learning for either modeling data directly (i.e., modeling observations 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

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

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:

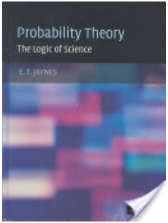
$$p(\theta|\mathbf{X}) = \frac{p(\mathbf{X}|\theta)p(\theta)}{p(\mathbf{X})} = \frac{p(\mathbf{X}|\theta)p(\theta)}{\int_\theta p(\mathbf{X}|\theta)p(\theta)d\theta}$$

In words:

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

Ignorance priors

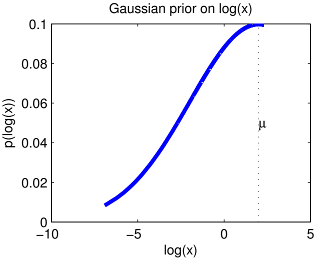
- Sometimes we don't have previous observations to formulate priors.
- Jaynes suggests using a maximum entropy prior.
- An ignorance prior is a prior probability distribution where equal probability is assigned to all possibilities.
- Ignorance priors can be motivated via invariance/symmetry (transformation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

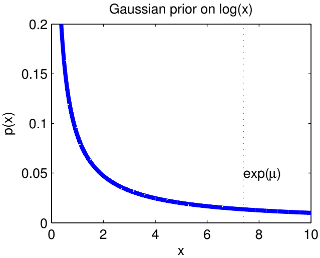


Jaynes, Edwin T. *Probability theory: the logic of science*. Cambridge university press, 2003.

Priors for positive values

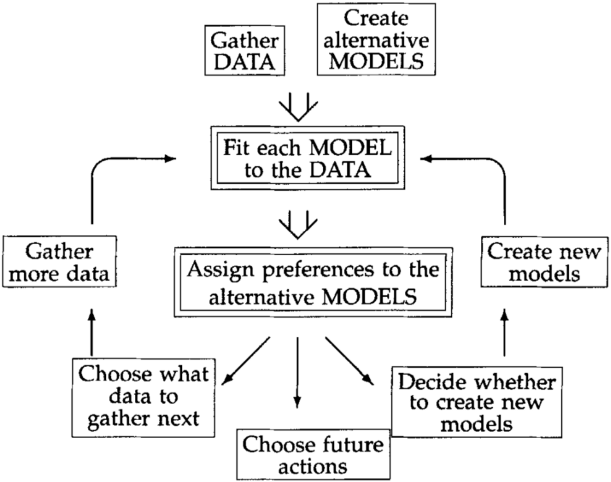
- Some things can not be less than zero.
    - Counts of observed photons.
    - Multiplicative "bias" fields.
    - Lengths, areas, volumes, etc.
- Formulate the model via logarithms, and impose a prior on these.





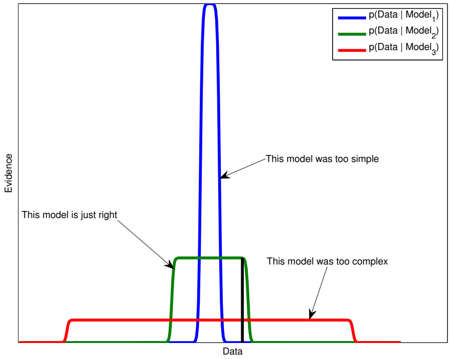
Jeffreys, Harold. "*An invariant form for the prior probability in estimation problems.*" In Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, vol. 186, no. 1007, pp. 453-461. The Royal Society, 1946.

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)

$$p(\mathbf{x}|\mathcal{M}) = \int_\theta p(\mathbf{x},\theta|\mathcal{M})d\theta$$

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 one of determining $p(\mathbf{y}^*|\mathbf{x}^*, \mathbf{Y}, \mathbf{X})$.

General aim of medical image computing

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

$$p(\mathbf{y}^*|\mathbf{x}^*,\mathbf{Y},\mathbf{X},\mathcal{M}) = \int_\theta p(\mathbf{y}^*,\theta|\mathbf{x}^*,\mathbf{Y},\mathbf{X},\mathcal{M})d\theta$$
$$= \int_\theta p(\mathbf{y}^*|\mathbf{x}^*,\mathbf{Y},\mathbf{X},\theta,\mathcal{M})p(\theta|\mathcal{M})d\theta$$

Predictions may also be made by averaging over models.

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

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.

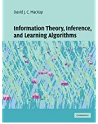
$$\hat{\theta} = \arg\max_{\theta} \log p(\mathbf{X}, \theta)$$

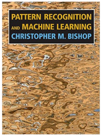
- Model selection via cross-validation.

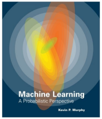
Useful Textbooks

The following books are suggested for all things Bayesian...

- MacKay, David JC. *Information theory, inference and learning algorithms*. Cambridge University Press, 2003. `http://www.inference.org.uk/itprnn/book.html`
- Bishop, Christopher M. *Pattern recognition and machine learning*. Springer, 2006.
- Murphy, Kevin P. *Machine learning: a probabilistic perspective*. MIT Press, 2012.







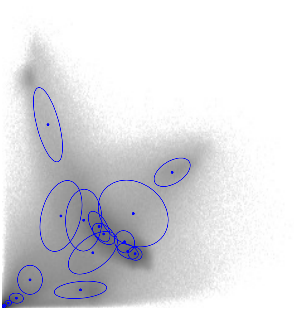
1. Introduction

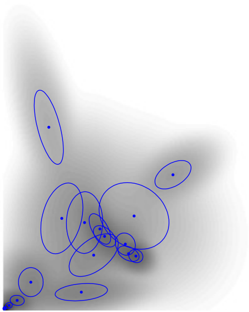
2. Segmentation
   - Mixture of Gausians
   - "Bias" correction
   - Deformable tissue priors

3. Longitudinal Registration

Mixture of Gausians

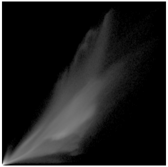
$$\mathcal{E} = -\log p(\mathbf{f}|\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\gamma})$$
$$= -\sum_{i=1}^{I} \log \left(\sum_{k=1}^{K} \frac{\gamma_k}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{(f_i - \mu_k)^2}{2\sigma_k^2} \right) \right)$$



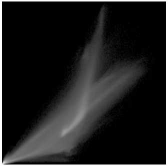


Incorporating "bias" correction

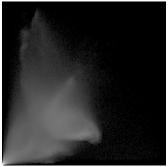
$$\mathcal{E} = -\sum_{i=1}^{I} \log\left(\sum_{k=1}^{K} \frac{\gamma_k}{\sqrt{2\pi \frac{\sigma_k^2}{\rho_i(\boldsymbol{\beta})^2}}} \exp\left(-\frac{\left(f_i - \frac{\mu_k}{\rho_i(\boldsymbol{\beta})}\right)^2}{2\frac{\sigma_k^2}{\rho_i(\boldsymbol{\beta})^2}}\right)\right)$$



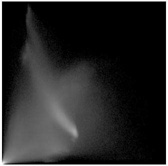
Original



Corrected



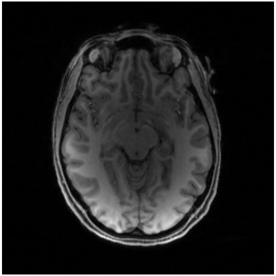
Original



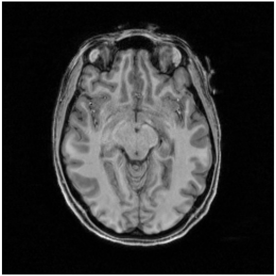
Corrected

Incorporating "bias" correction

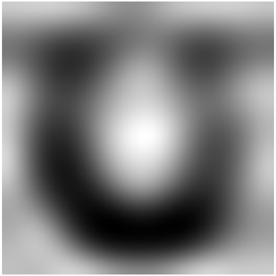
$$\mathcal{E} = -\sum_{i=1}^{I} \log \left(\rho_i(\boldsymbol{\beta}) \sum_{k=1}^{K} \frac{\gamma_k}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{(\rho_i(\boldsymbol{\beta}) f_i - \mu_k)^2}{2\sigma_k^2} \right) \right)$$



Original



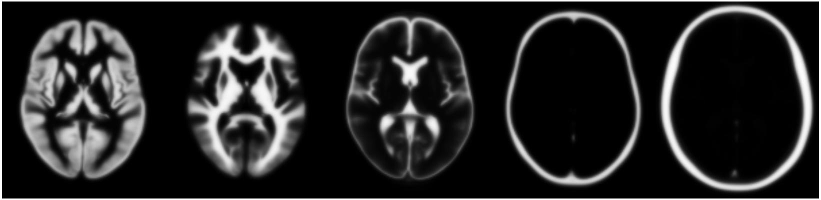
Corrected



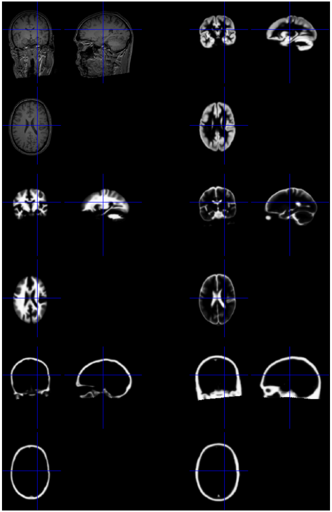
Field

Incorporating deformable tissue priors

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

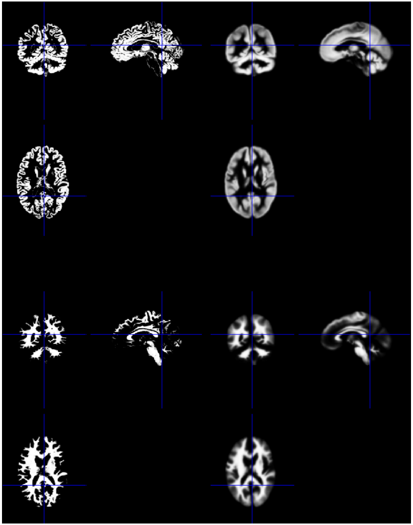


Incorporating deformable tissue priors



Registration achieved by optimising objective function w.r.t. $\boldsymbol{\alpha}$.
$b_{ik}(\boldsymbol{\alpha})$ denotes tissue probability of class k at voxel i, after warping by parameters $\boldsymbol{\alpha}$.

Latent variables



Optimisation done via EM.

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

$$
\begin{aligned}
p(\mathbf{f}, \boldsymbol{\theta}) &= \int_{\mathbf{z}} p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta}) dz \\
\text{where} & \\
\boldsymbol{\theta} &= \{\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\alpha}\}
\end{aligned}
$$

Extensions

- Intensity distributions of tissue classes are estimated afresh each time.
  - Intensity priors can be used to inform their estimation, using *Variational Bayes*.
  - These can be learned from a population of images.
- Spatial tissue priors can be learned from a population of brain scans.
  - Can introduce a semi-supervised learning scheme.

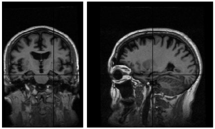
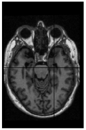
- Ashburner, John, and Karl J. Friston. "*Unified segmentation.*" Neuroimage 26, no. 3 (2005): 839-851.
- `http://www.fil.ion.ucl.ac.uk/spm/software/spm12/`, `spm12/spm_preproc_run.m`.
- Blaiotta, Claudia, M. Jorge Cardoso, and John Ashburner. "*Variational inference for medical image segmentation.*" Computer Vision and Image Understanding 151 (2016): 14-28.

1 Introduction

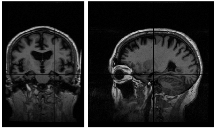
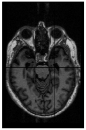
2 Segmentation

3 Longitudinal Registration
- Differential Bias Fields
- Rigid-Body
- Diffeomorphisms
- Combined Model

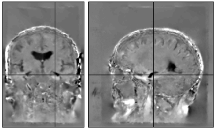
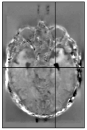
Longitudinal data: OAS2_0002

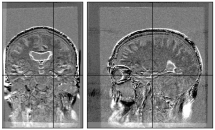
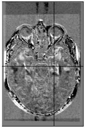
First scan (0 days) after rigid body registration.

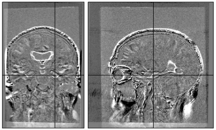
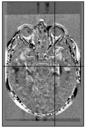
Second scan (1895 days) after rigid body registration.

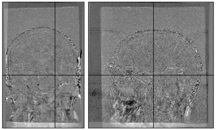
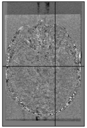
Jacobian determinants windowed between 0.7 and 1.3.

Residuals from rigid body registration, without bias correction.

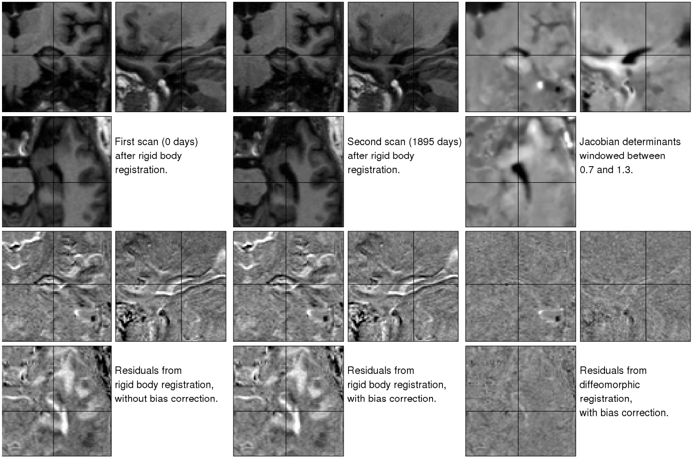



Residuals from rigid body registration, with bias correction.

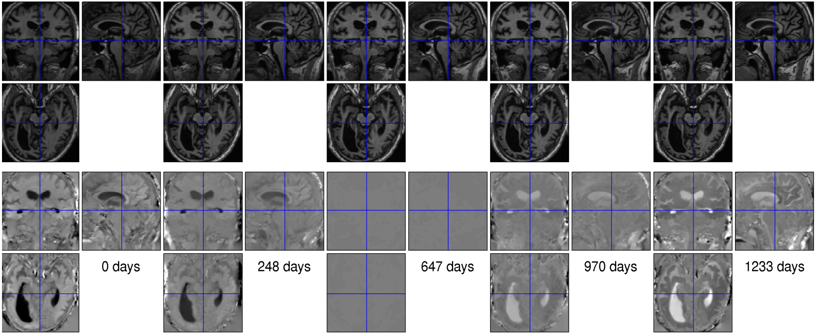



Residuals from diffeomorphic registration, with bias correction.

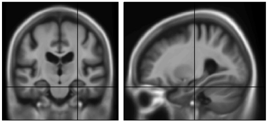
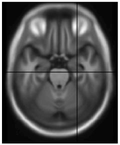
Longitudinal data: OAS2_0002



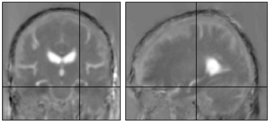
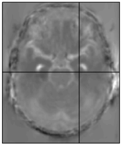
Longitudinal data: OAS2_0048



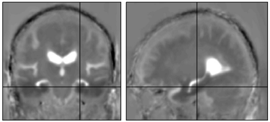
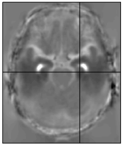
Longitudinal data: Averages

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:

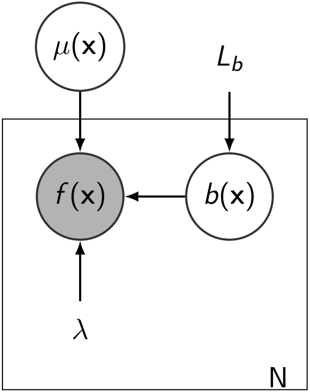
$$\mathcal{E}(\boldsymbol{\theta}) = \sum_i r_i^2(\boldsymbol{\theta})$$

- 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).

Differential 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

Differential 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

Differential 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.

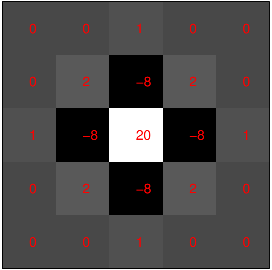
$$\exp(b) = \lim_{n \to \infty} \left(1 + \frac{b}{n}\right)^n$$

Differential Bias Fields: Regularisation

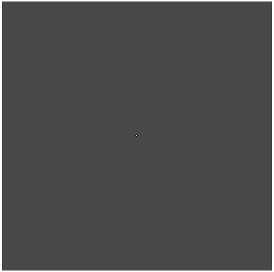
Penalise sum of squares of second derivatives:

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

Differential operator ($L_b^\dagger L_b$)

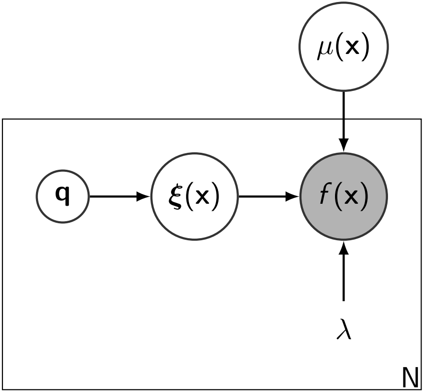


Differential operator (zoomed out)



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 ($\mathbf{R_q} \in SE(3)$) is computed via a matrix exponential:

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

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.

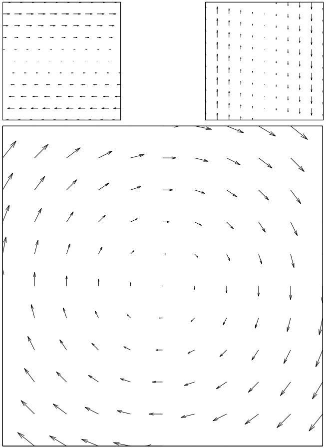
Rigid-Body: Exponential Map

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

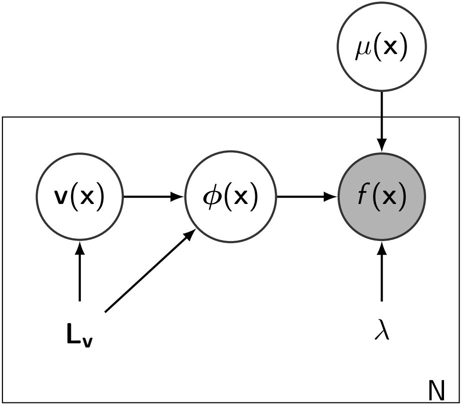
$$\mathbf{R_q} = \exp \begin{bmatrix} 0 & q_1 \\ -q_1 & 0 \end{bmatrix}$$

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

N – number of images

Note: Diffeomorphic deformations are computed via a Riemannian exponential.

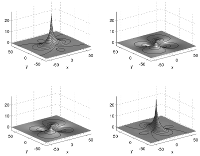
Diffeomorphisms: Exponential Map

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:

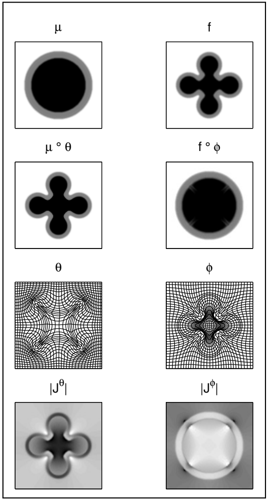
$$\dot{\phi}_{\mathbf{v}} = \left(\mathbf{K}_{\mathbf{v}}\left(\left|\mathbf{D}\phi_{\mathbf{v}}^{-1}\right|\left(\mathbf{D}\phi_{\mathbf{v}}^{-1}\right)^{T}\left(\mathbf{u}\circ\phi_{\mathbf{v}}^{-1}\right)\right)\right)\circ\phi_{\mathbf{v}}$$

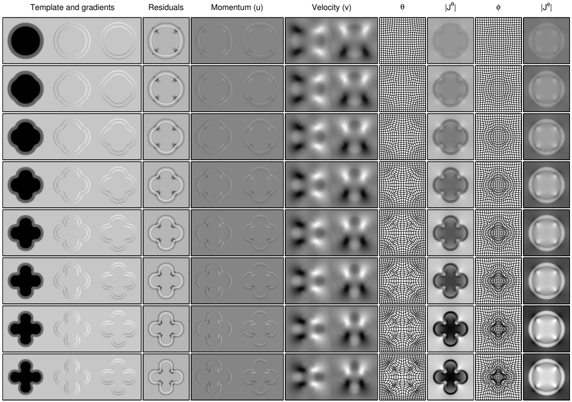


$\mathbf{K}_{\mathbf{v}}$ is the Green's function of $\mathbf{L}_{\mathbf{v}}^{\dagger}\mathbf{L}_{\mathbf{v}}$, such that:

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

Diffeomorphisms: Exponential Map





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

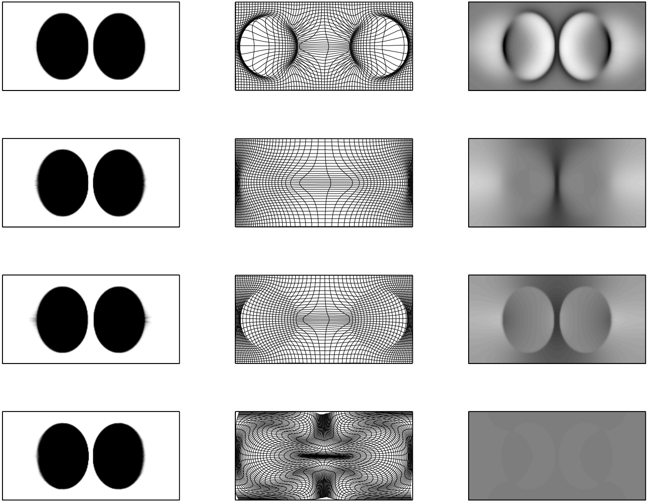
Diffeomorphisms: Regularisation

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

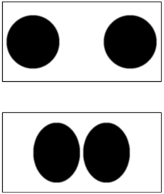
Three hyper-parameters are involved:

- 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.

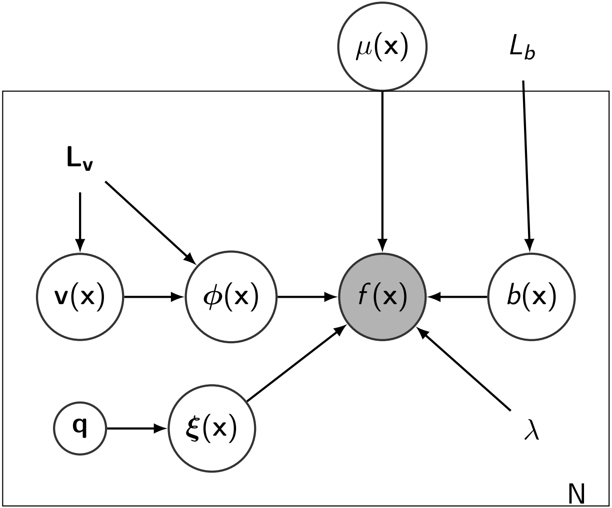
Diffeomorphisms: Regularisation



Two simulated images



Combined Model: Generative Model



- $f(\mathrm{x})$ – image
- μ – mean image
- λ – noise precision
- $b(\mathrm{x})$ – "bias" field
- L_b – bias field regularisation
- $\xi(\mathrm{x})$ – 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:

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

where:

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

"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`.


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)