Principles of DCM

Will Penny

26th May 2011

Dynamic Causal Modelling

Dynamic Causal Modelling is a framework studying large scale brain connectivity by fitting differential equation models to brain imaging data.

DCMs differ in their level of biological plausibility and the data features which they explain.

What is common to all DCMs are the Bayesian methods for parameter estimation and model comparison.

SPM currently includes

- DCM for event related potentials - Jean
- DCM for steady state responses - Rosalyn
- DCM for induced responses - Bernadette
- DCM for phase coupling - Bernadette

Dynamic Causal Modelling

SPM currently includes

- DCM for event related potentials - neural mass models fitted to sensor space ERPs
- DCM for steady state responses - linearised neural mass models fitted to source space spectra and cross spectra
- DCM for induced responses - bilinear differential equation models of spectrograms
- DCM for phase coupling - weakly coupled oscillator models of source space phase time series

Differential Equations

Differential equations have a long history in neuroscience from the Hodgkin-Huxley (1952) model of a spiking neuron, to Wilson-Cowan (1972) models of neural populations, to spatially distributed models of neural fields (Deco et al 2008).

They take the general form

$$\frac{dx}{dt} = f(x, u, w)$$

where x are neural state variables, u are experimental perturbations and w are biophysical parameters.

If the inputs and parameters are known these equations can be integrated eg by Euler method

$$x_{n+1} = x_n + f(x_n, u_n, w)\Delta_t$$

where Δ_t is the time step, to produce time series $x(t)$.

Forward Model

Neural state variables, x, are related to quantities derived from brain imaging data, y, by an observation equation

$$y = g(x) + e$$

where e is Gaussian observation noise. If the data features are in sensor space then this will include the lead field.

The differential equations and observation equation together describe a forward model. This specifices the likelihood $p(y|w)$.

A prior distribution over parameters $p(w)$, reflecting either ignorance or physiological constraints, is then updated to a posterior density $p(y|w)$ using an approximate Bayesian inference procedure called Variational Laplace (VL).

Questions

Questions I will address include

- What are generic properties of differential equations ?
- What is a Neural Mass model ?
- How does Variational Laplace work ?

Exponentials

We use the following shorthand for a time derivative

$$\dot{x} = \frac{dx}{dt}$$

The exponential function $x = \exp(t)$ is invariant to differentiation. Hence

$$\dot{x} = \exp(t)$$

and

$$\dot{x} = x$$

Hence $\exp(t)$ is the solution of the above differential equation.

Initial Values and Fixed Points

An exponential increase ($a > 0$) or decrease ($a < 0$) from initial condition x_0

$$x = x_0 \exp(at)$$

has derivative

$$\dot{x} = ax_0 \exp(at)$$

The top equation is therefore the solution of the differential equation

$$\dot{x} = ax$$

with initial condition x_0.

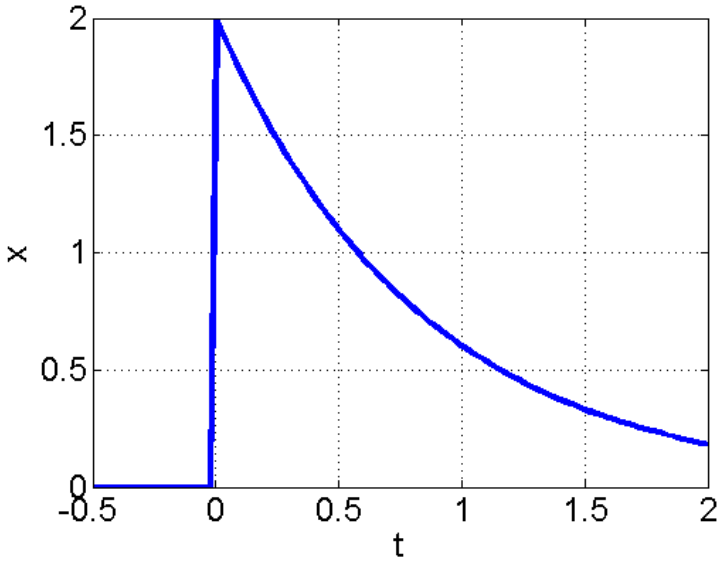
The values of x for which $\dot{x} = 0$ are referred to as Fixed Points (FPs). For the above the only fixed point is at $x = 0$.

Time constants

The figure shows

$$\dot{x} = ax$$

with $a = -1.2$ and intial value $x_0 = 2$.



The time constant is $\tau = -1/a$.

The time at which x decays to half its initial value is

$$\tau_h = \frac{1}{a} \log(1/2)$$

which equals $\tau_h = 0.58$.

Matrix Exponential

If x is a vector whose evolution is governed by a system of linear differential equations we can write

$$\dot{x} = Ax$$

where A describes the linear dependencies.

The only fixed point is at $x = 0$.

For initial conditions x_0 the above system has solution

$$x_t = \exp(At)x_0$$

where $\exp(At)$ is the matrix exponential (written expm in matlab) (Moler and Van Loan, 2003).

Eigendecomposition

The equation

$$\dot{x} = Ax$$

can be understood by representing A with an eigendecomposition, with eigenvalues λ_k and eigenvectors q_k that satisfy (Strang, p. 255)

$$A = Q\Lambda Q^{-1}$$

We can then use the identity

$$\exp(A) = Q\exp(\Lambda)Q^{-1}$$

Because Λ is diagonal, the matrix exponential simplifies to a simple exponential function over each diagonal element.

Dynamical Modes

This tells us that the original dynamics

$$\dot{x} = Ax$$

has a solution

$$x_t = \exp(At)$$

that can be represented as a linear sum of k independent dynamical modes

$$x_t = \sum_k q_k \exp(\lambda_k t)$$

where q_k and λ_k are the kth eigenvector and eigenvalue of A. For $\lambda_k > 0$ we have an unstable mode.

For $\lambda_k < 0$ we have a stable mode, and the magnitude of λ_k determines the time constant of decay to the fixed point.

The eigenvalues can also be complex. This gives rise to oscillations.

Spiral

A spiral occurs in a two-dimensional system when both eigenvalues are a complex conjugate pair. For example (Wilson, 1999)

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} -2 & -16 \\ 4 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

has

$$\begin{aligned} \lambda_1 &= -2 + 8i \\ \lambda_2 &= -2 - 8i \end{aligned}$$

giving solutions (for initial conditions $x = [1, 1]^T$)

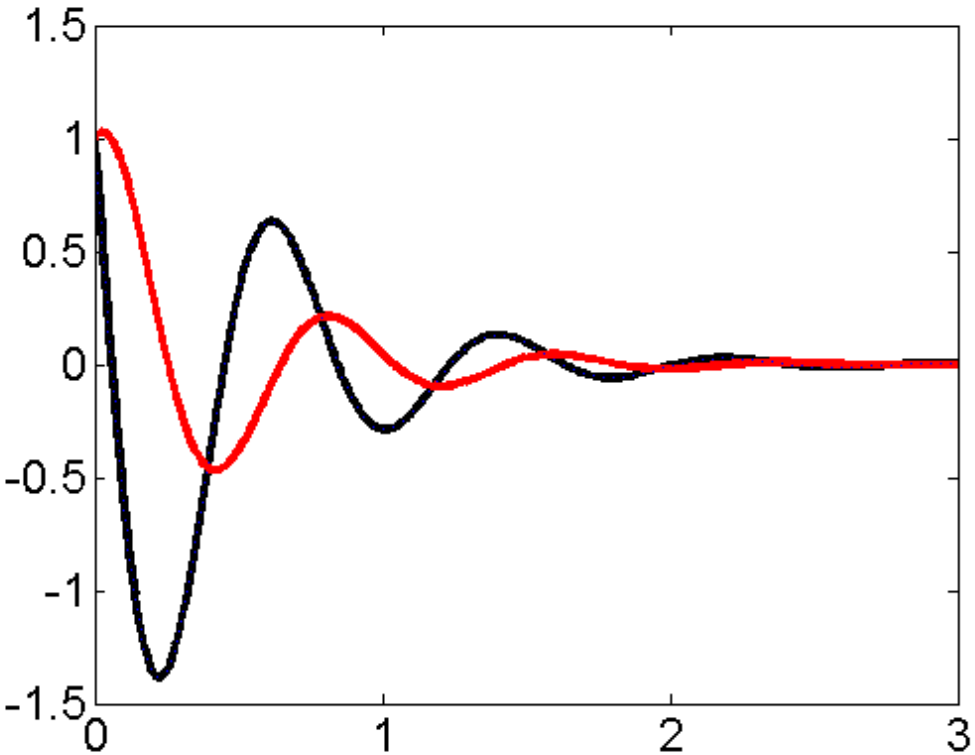
$$\begin{aligned} x_1(t) &= \exp(-2t)\left[\cos(8t) - 2\sin(8t)\right] \\ x_2(t) &= \exp(-2t)\left[\cos(8t) + 0.5\sin(8t)\right] \end{aligned}$$

Spiral

We plot time series solutions

$$
\begin{aligned}
x_1(t) &= \exp(-2t)\left(\cos(8t) - 2\sin(8t)\right) \\
x_2(t) &= \exp(-2t)\left(\cos(8t) + 0.5\sin(8t)\right)
\end{aligned}
$$

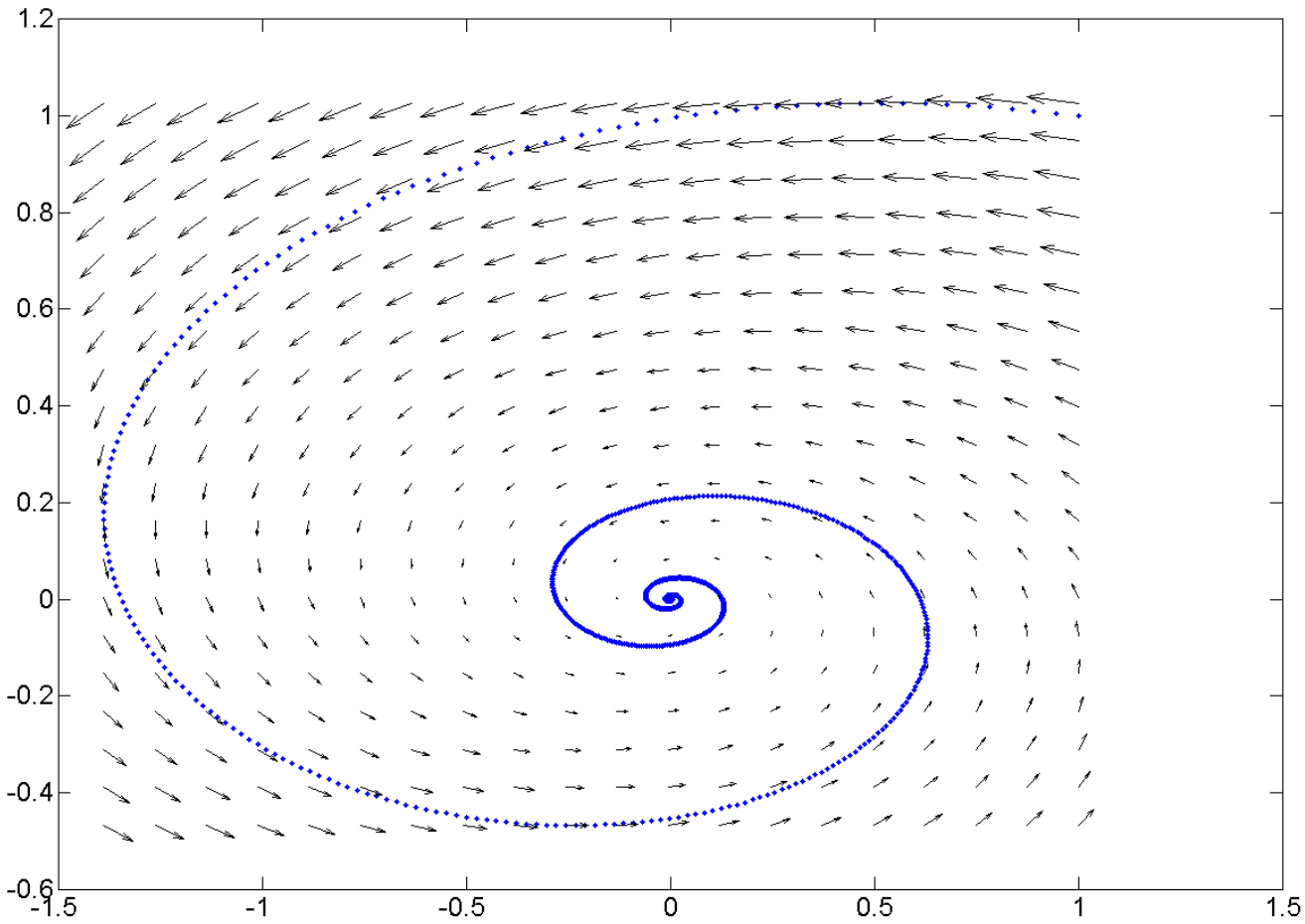
for x_1 (black) and x_2 (red).



8 radians = 1.3Hz

Spiral State Space

Plotting x_2 against x_1 gives the state-space representation.



x_1

Embedding

Univariate higher order differential equations can be represented as multivariate first order DEs.

For example

$$\ddot{v} = \frac{H}{\tau} u_t - \frac{2}{\tau} \dot{v} - \frac{1}{\tau^2} v$$

can be written as

$$\begin{aligned} \dot{v} &= c \\ \dot{c} &= \frac{H}{\tau} u_t - \frac{2}{\tau} c - \frac{1}{\tau^2} v \end{aligned}$$

Kernels

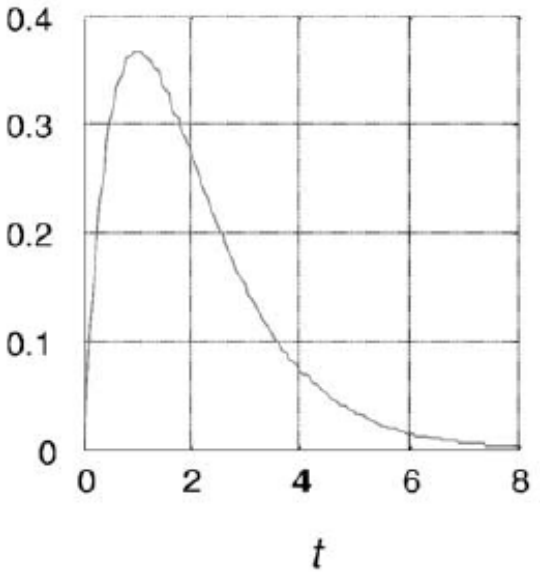
The previous differential equation has a solution given by the integral

$$v(t) = \int u(t)h(t - t')dt'$$

where

$$h(t) = \frac{H}{\tau} t \exp(-t/\tau)$$

is a kernel. In this case it is an alpha function synapse with magnitude H and time constant τ



Convolution

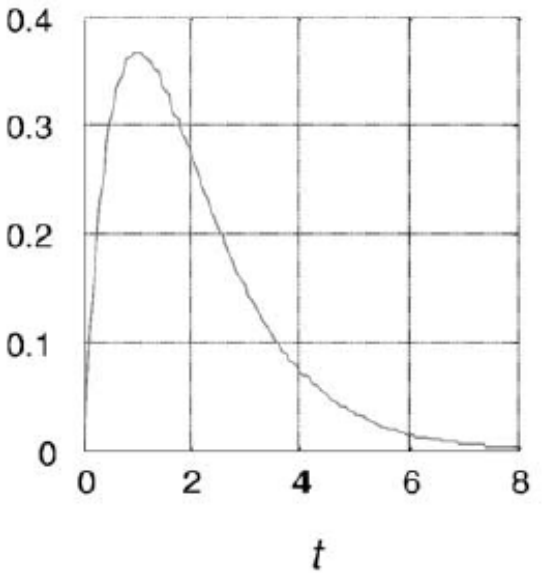
The previous integral can be written as

$$v = u \otimes h$$

where $\otimes$ is the convolution operator.

$$h(t) = \frac{H}{\tau} t \exp(-t/\tau)$$

is a kernel. In this case it is an alpha function synapse with magnitude H and time constant τ



Neural Mass Model

Jansen and Rit (1995), building on the work of Lopes Da Sliva and others, developed a biologically inspired model of EEG activity. It was originally developed to explain alpha activity and Event-Related Potentials (ERPs).

It models a cortical unit with three subpopulations of cells

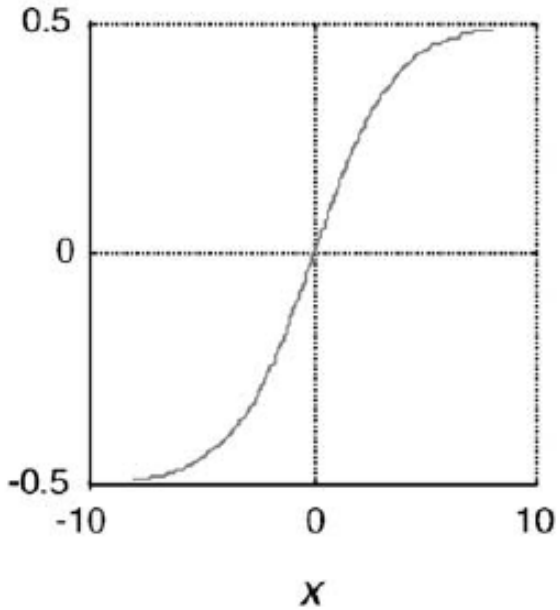
- Stellate cells with average membrane potential v_s.
- Pyramidal cells with average membrane potential v_p.
- Inhibitory interneurons with average membrane potential v_i.

Here I describe the Neural Mass model for a single cortical unit as formulated in David et al. (2006).

Firing Rate Curves

Membrane potentials, x, are transformed into firing rates via sigmoidal functions

$$s(x) = \frac{1}{1 + \exp(-rx)} - \frac{1}{2}$$

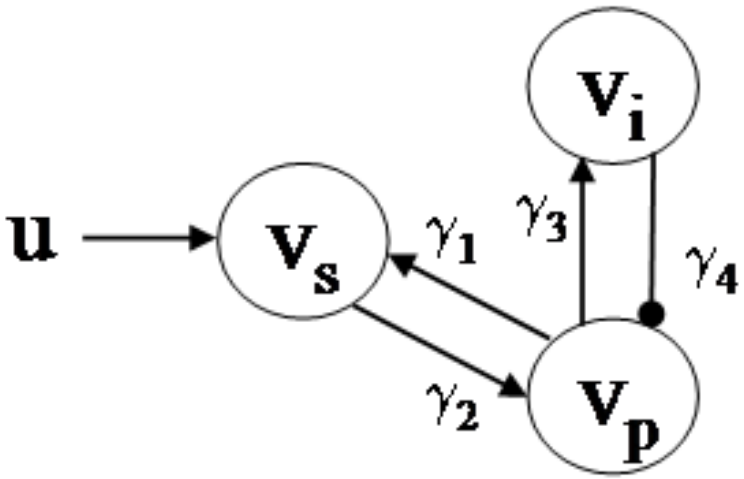


Negative firing rates here allow systems to have a stable fixed point at $x = 0$. All firing rates are therefore considered as deviation from steady state values.

Inhibitory Interneurons

The inhibitory interneurons receive excitatory input from the pyramidal cells

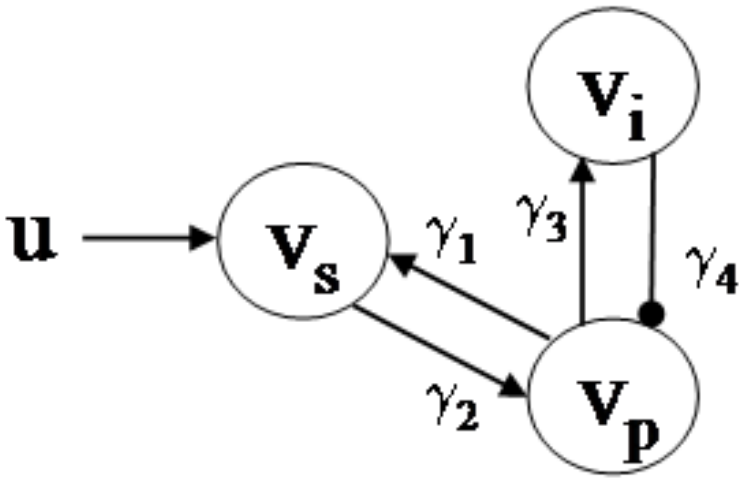
$$v_i = \gamma_3 s(v_p) \otimes h_e$$



Stellate Cells

The stellate cells receive external input from thalamus or other cortical regions and excitatory feedback from pyramidal cells

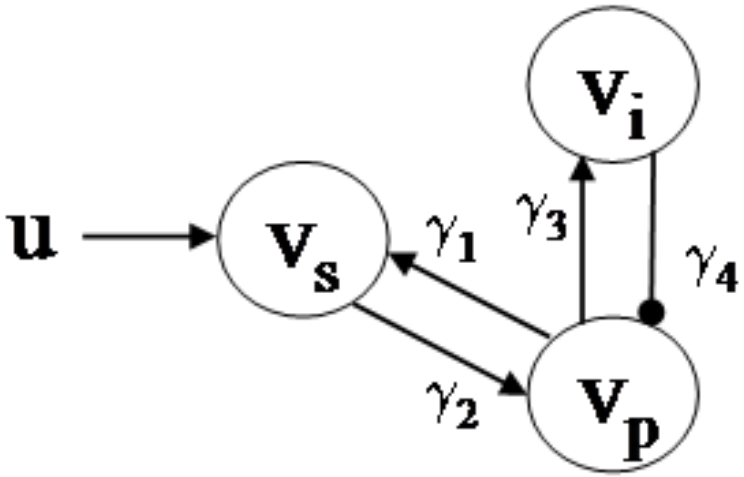
$$v_s = (s(u) + \gamma_1 s(v_p)) \otimes h_e$$



Pyramidal Cells

The pyramidal cells receive excitatory input from stellate cells and inhibitory input from interneurons. This produces both excitatory v_{pe} and inhibitory v_{pi} postsynaptic potentials. This formulation is due to David et al (2006).

$$
\begin{aligned}
v_{pe} &= \gamma_2 s(v_s) \otimes h_e \\
v_{pi} &= \gamma_4 s(v_i) \otimes h_i \\
v_p &= v_{pe} - v_{pi}
\end{aligned}
$$



Differential Equations

The integral equations become

$$
\begin{aligned}
v_i &= \gamma_3 s(v_p) \otimes h_e \\
v_s &= \left(s(u) + \gamma_1 s(v_p)\right) \otimes h_e \\
v_{pe} &= \gamma_2 s(v_s) \otimes h_e \\
v_{pi} &= \gamma_4 s(v_i) \otimes h_i \\
v_p &= v_{pe} - v_{pi}
\end{aligned}
$$

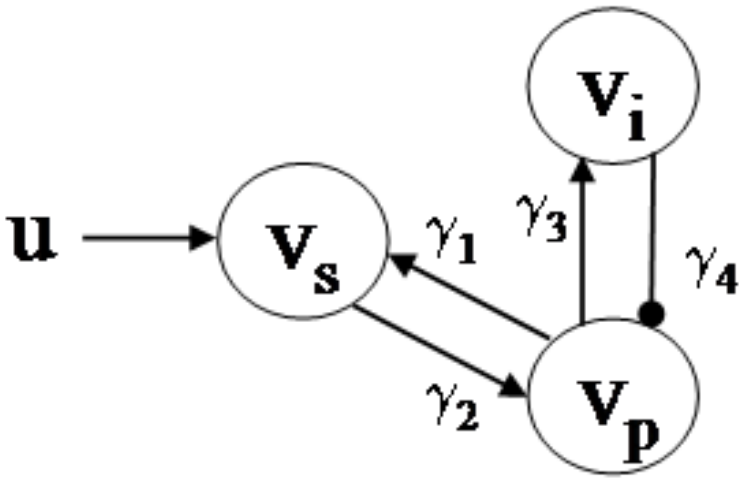
the differential equations

$$
\begin{aligned}
\dot{v}_i &= c_i \\
\dot{c}_i &= \frac{H_e}{\tau_e}\gamma_3 s(v_p(t)) - \frac{2}{\tau_e}c_i - \frac{1}{\tau_e^2}v_i \\
\dot{v}_s &= c_s \\
\dot{c}_s &= \frac{H_e}{\tau_e}\gamma_3(s(u(t)) + \gamma_1 s(v_p(t)) - \frac{2}{\tau_e}c_s - \frac{1}{\tau_e^2}v_s \\
\dot{v}_{pe} &= c_{pe} \\
\dot{c}_{pe} &= \frac{H_e}{\tau_e}\gamma_2 s(v_s(t)) - \frac{2}{\tau_e}c_{pe} - \frac{1}{\tau_e^2}v_{pe} \\
\dot{v}_{pi} &= c_{pi} \\
\dot{c}_{pi} &= \frac{H_i}{\tau_i}\gamma_4 s(v_i(t)) - \frac{2}{\tau_i}c_{pi} - \frac{1}{\tau_i^2}v_{pi} \\
\dot{v}_p &= c_{pe} - c_{pi}
\end{aligned}
$$

Intrinsic Connectivity

Based on the relative counts of numbers of synapses in cat and mouse visual and somato-sensory cortex Jansen and Rit (1995) determined the following connectivity values.

$$
\begin{aligned}
\gamma_1 &= C \\
\gamma_2 &= 0.8C \\
\gamma_3 &= 0.25C \\
\gamma_4 &= 0.25C
\end{aligned}
$$



Nonlinear Regression

We consider the framework implemented in the SPM function *spm-nlsi-GN.m*. It implements Bayesian estimation of nonlinear models of the form

$$y = g(w) + e$$

where $g(w)$ is some nonlinear function of parameters w, and e is zero mean additive Gaussian noise with covariance C_y. The likelihood of the data is therefore

$$p(y|w, \lambda) = \mathrm{N}(y; g(w), C_y)$$

where N denotes a multivariate normal density.

Nonlinear Regression

The precision is the inverse of the variance.

The error *precision* matrix is assumed to decompose linearly

$$C_y^{-1} = \sum_i \exp(\lambda_i) Q_i$$

where Q_i are known precision basis functions and λ are hyperparameters eg $Q = I$, noise precision $s = \exp(\lambda)$.

Priors

We allow Gaussian priors over model parameters

$$p(w) = \mathrm{N}(w; \mu_w, C_w)$$

where the prior mean and covariance are assumed known.

The hyperparameters are constrained by the prior

$$p(\lambda) = \mathrm{N}(\lambda; \mu_\lambda, C_\lambda)$$

This is not Empirical Bayes.

Joint Log Likelihood

The above distributions allow one to write down an expression for the joint log likelihood of the data, parameters and hyperparameters

$$L(w,\lambda) = \log[p(y|w,\lambda)p(w)p(\lambda)]$$

Here it splits into three terms

$$\begin{aligned} L(w,\lambda) &= \log p(y|w,\lambda) \\ &+ \log p(w) \\ &+ \log p(\lambda) \end{aligned}$$

Joint Log Likelihood

The joint log likelihood is composed of sum squared precision weighted prediction errors and entropy terms

$$\begin{aligned} L &= -\frac{1}{2} e_y^T C_y^{-1} e_y - \frac{1}{2} \log |C_y| - \frac{N_y}{2} \log 2\pi \\ &- \frac{1}{2} e_w^T C_w^{-1} e_w - \frac{1}{2} \log |C_w| - \frac{N_w}{2} \log 2\pi \\ &- \frac{1}{2} e_\lambda^T C_\lambda^{-1} e_\lambda - \frac{1}{2} \log |C_\lambda| - \frac{N_\lambda}{2} \log 2\pi \end{aligned}$$

where prediction errors are the difference between what is expected and what is observed

$$\begin{aligned} e_y &= y - g(m_w) \\ e_w &= m_w - \mu_w \\ e_\lambda &= m_\lambda - \mu_\lambda \end{aligned}$$

VL Posteriors

The Variational Laplace (VL) algorithm, implemented in *spm-nlsi-GN.m*, assumes an approximate posterior density of the following factorised form

$$\begin{aligned} q(w,\lambda|y) &= q(w|y)q(\lambda|y) \\ q(w|y) &= \mathrm{N}(w; m_w, S_w) \\ q(\lambda|y) &= \mathrm{N}(\lambda; m_\lambda, S_\lambda) \end{aligned}$$

This is a fixed-form variational method (Bishop, 2006).

Variational Energies

The approximate posteriors are estimated by minimising the Kullback-Liebler (KL) divergence between the true posterior and these approximate posteriors. This is implemented by maximising the following (negative) variational energies

$$I(w) = \int L(w,\lambda)q(\lambda)$$

$$I(\lambda) = \int L(w,\lambda)q(w)$$

Gradient Ascent

This maximisation is effected by first computing the gradient and curvature of the variational energies at the current parameter estimate, $m_w(old)$. For example, for the parameters we have

$$\begin{aligned} j_w(i) &= \frac{dI(w)}{dw(i)} \\ H_w(i,j) &= \frac{d^2 I(w)}{dw(i)dw(j)} \end{aligned}$$

where i and j index the ith and jth parameters, j_w is the gradient vector and H_w is the curvature matrix. The estimate for the posterior mean is then given by

$$m_w(new) = m_w(old) + \Delta m_w$$

Adaptive Step Size

The change is given by

$$\Delta m_w = -H_w^{-1} j_w$$

which is equivalent to a Newton update. This implements a step in the direction of the gradient with a step size given by the inverse curvature. Big steps are taken in regions where the gradient changes slowly (low curvature).

Adaptive Step Size

The change is given by

$$\Delta m_w = [\exp(vH_w) - I]\, H_w^{-1} j_w$$

This last expression implements a 'temporal regularisation' with parameter v (Friston et al. 2007). In the limit $v \to \infty$ the update reduces to

$$\Delta m_w = -H_w^{-1} j_w$$

which is equivalent to a Newton update. This implements a step in the direction of the gradient with a step size given by the inverse curvature. Big steps are taken in regions where the gradient changes slowly (low curvature).

Approach to Limit

State equation

$$\tau \dot{v} = V_0 + V_a - v$$

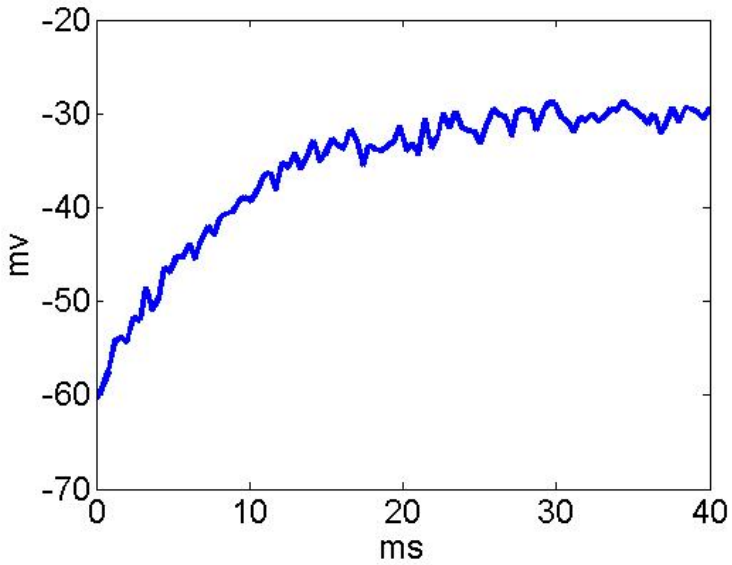
Observation equation

$$y = v + e$$

For this case we have an analytic solution

$$y(t) = V_0 + V_a[1 - \exp(-t/\tau)] + e(t)$$

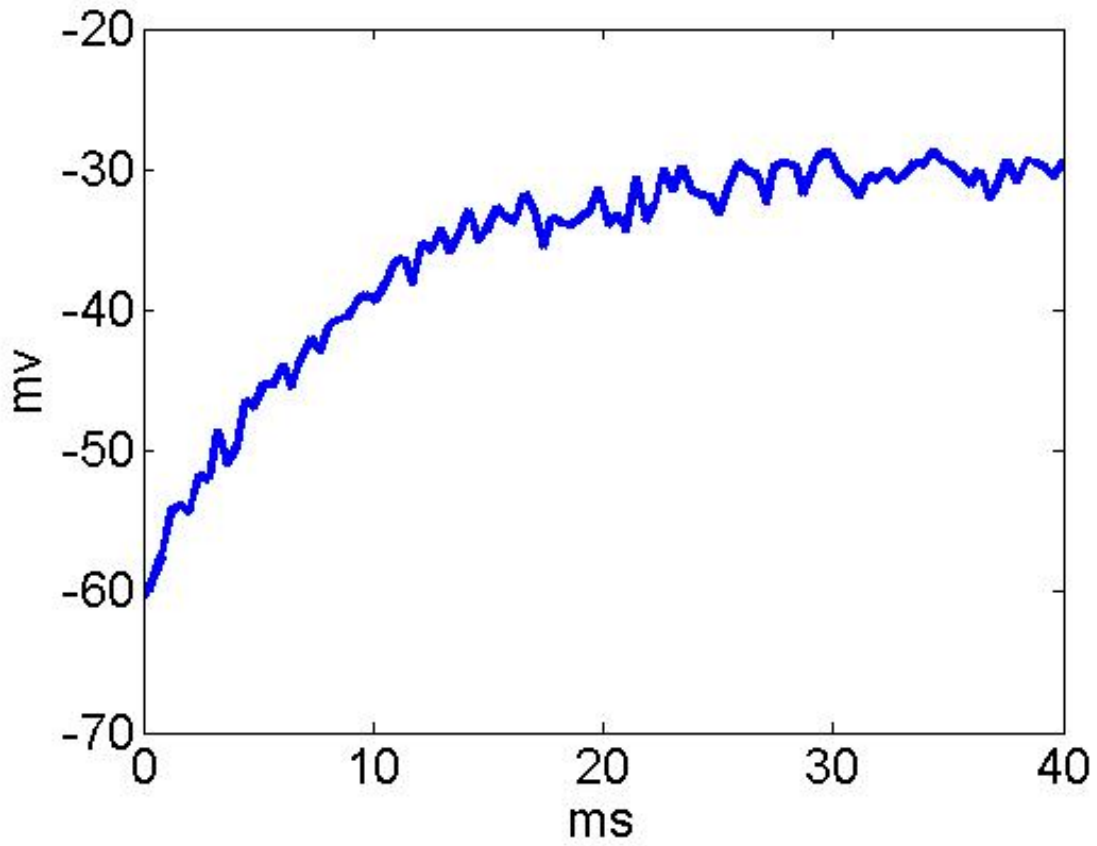
Initial value $v = V_0 = -60$.



Fixed point at $v = V_0 + V_a$ which is approached with time constant τ.

Approach to Limit

$$y(t) = -60 + V_a[1 - \exp(-t/\tau)] + e(t)$$



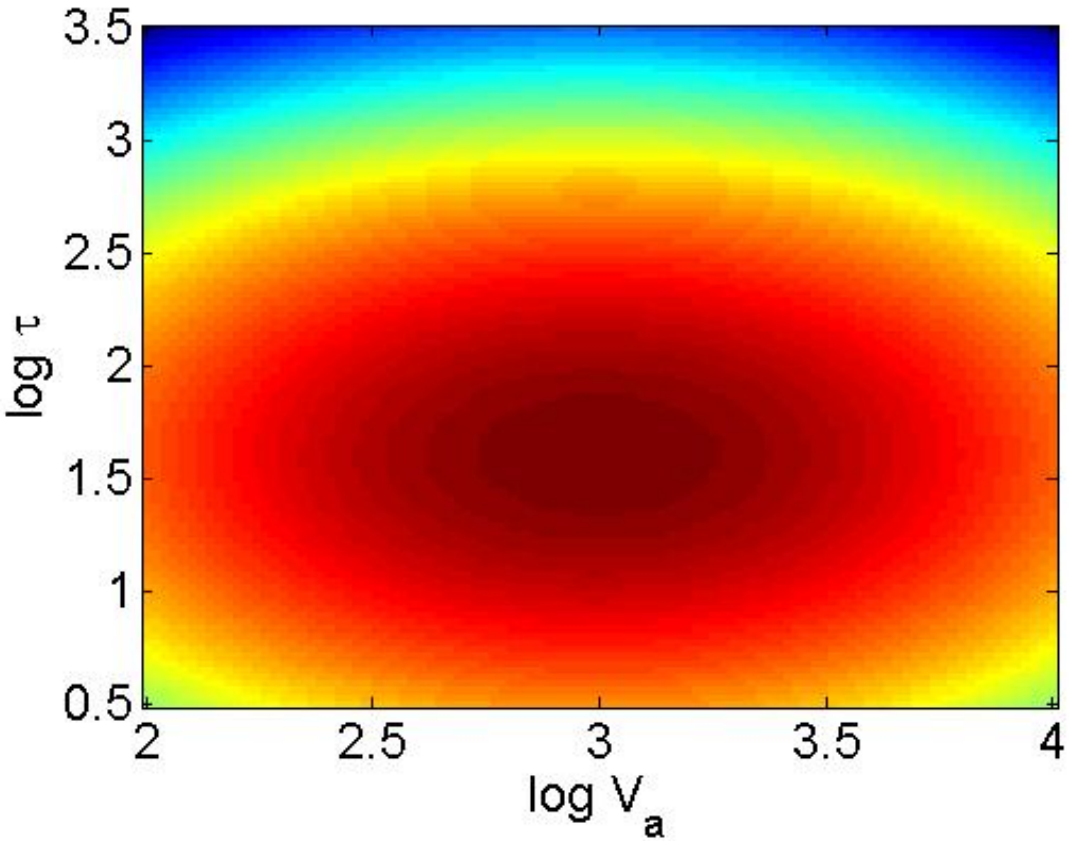
$$V_a = 30, \tau = 8$$

Noise precision

$$s = \exp(\lambda) = 1$$

Prior Landscape

A plot of $\log p(w)$ where $w = [\log \tau, \log V_a]$

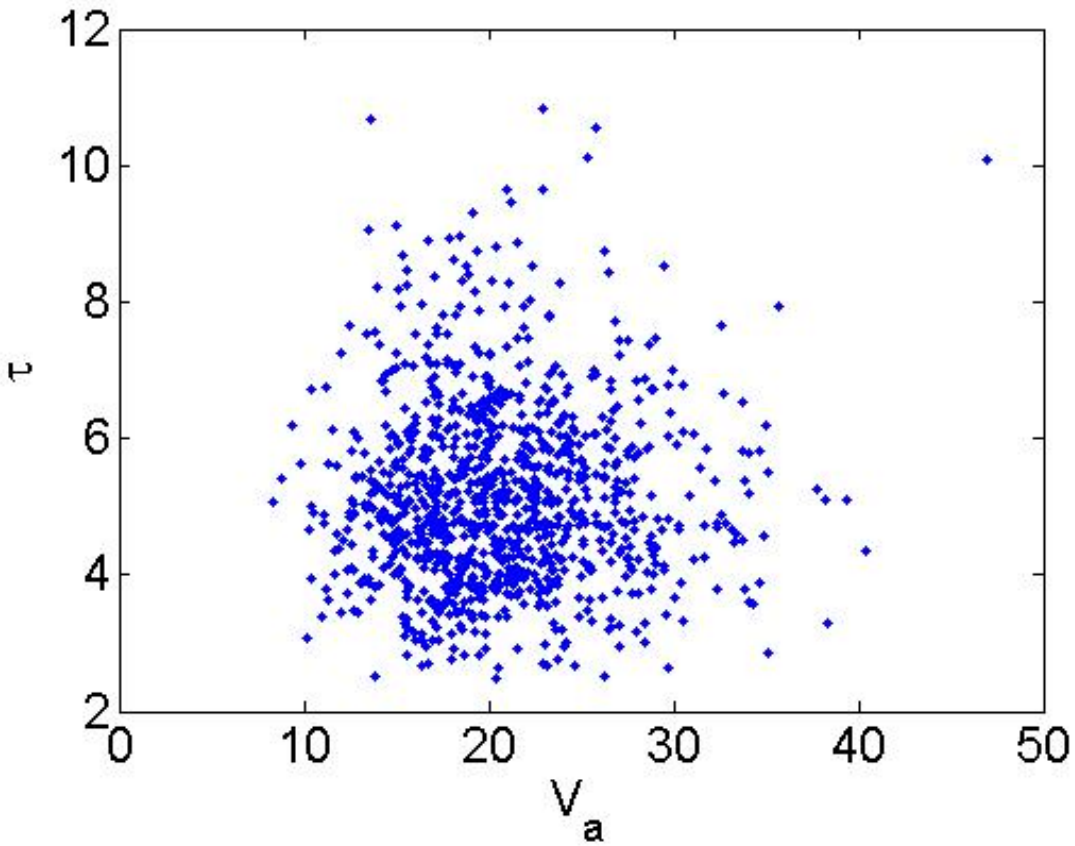


$$\mu_w = [3, 1.6]^T, C_w = diag([1/16, 1/16]);$$

Samples from Prior

The true model parameters are unlikely apriori

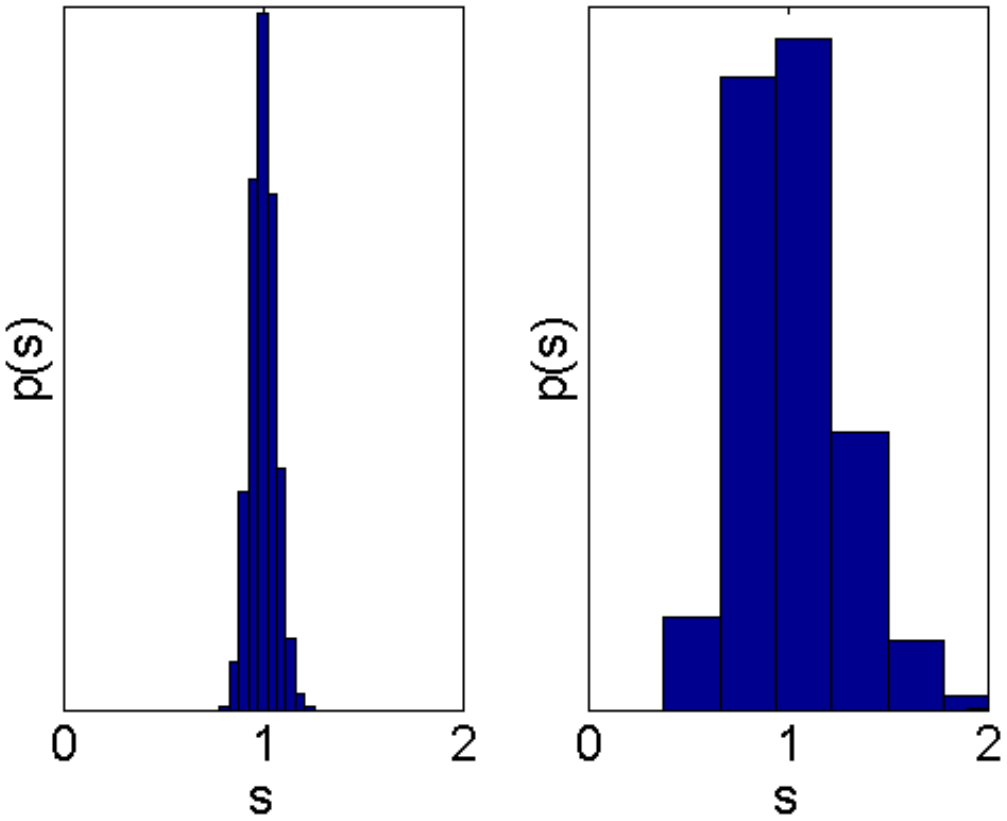
$$V_a = 30, \tau = 8$$



Prior Noise Precision

$Q = I$. Noise precision $s = \exp(\lambda)$ with

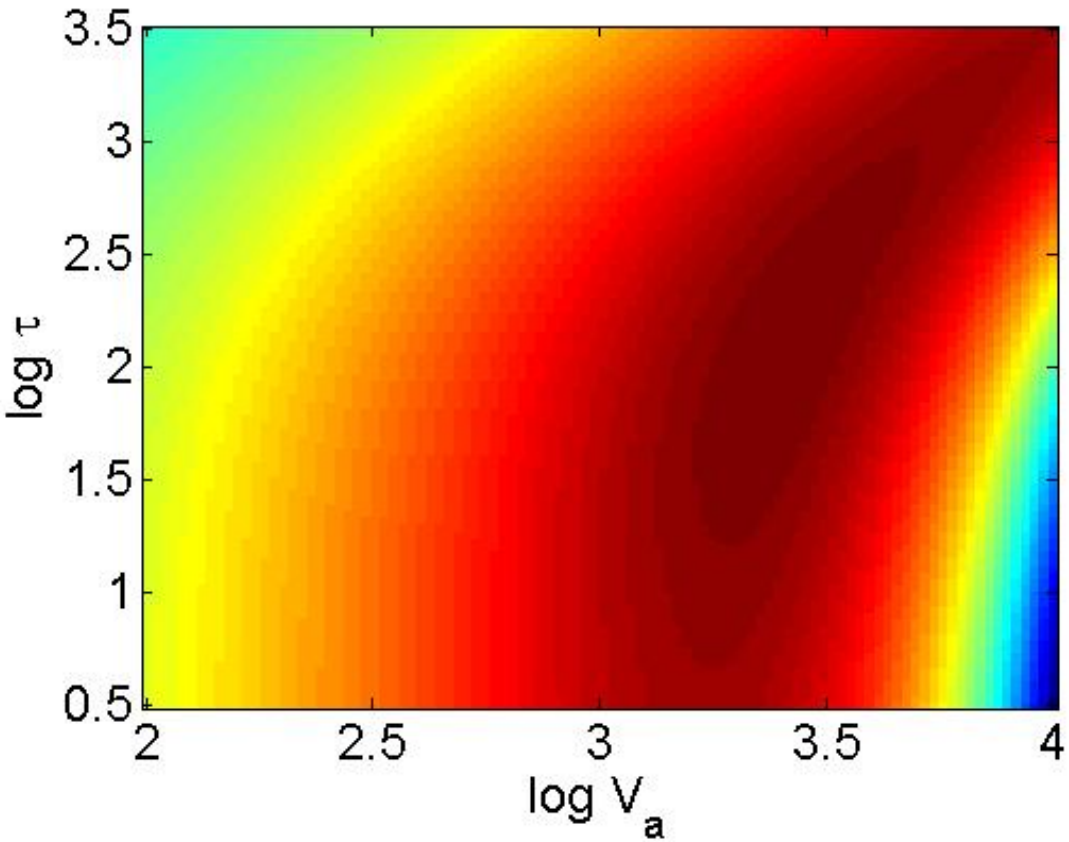
$$p(\lambda) = \mathrm{N}(\lambda; \mu_\lambda, C_\lambda)$$



with $\mu_\lambda = 0$. We used $C_\lambda = 1/16$ (left) and $C_\lambda = 1/4$ (right). True noise precision, $s = 1$.

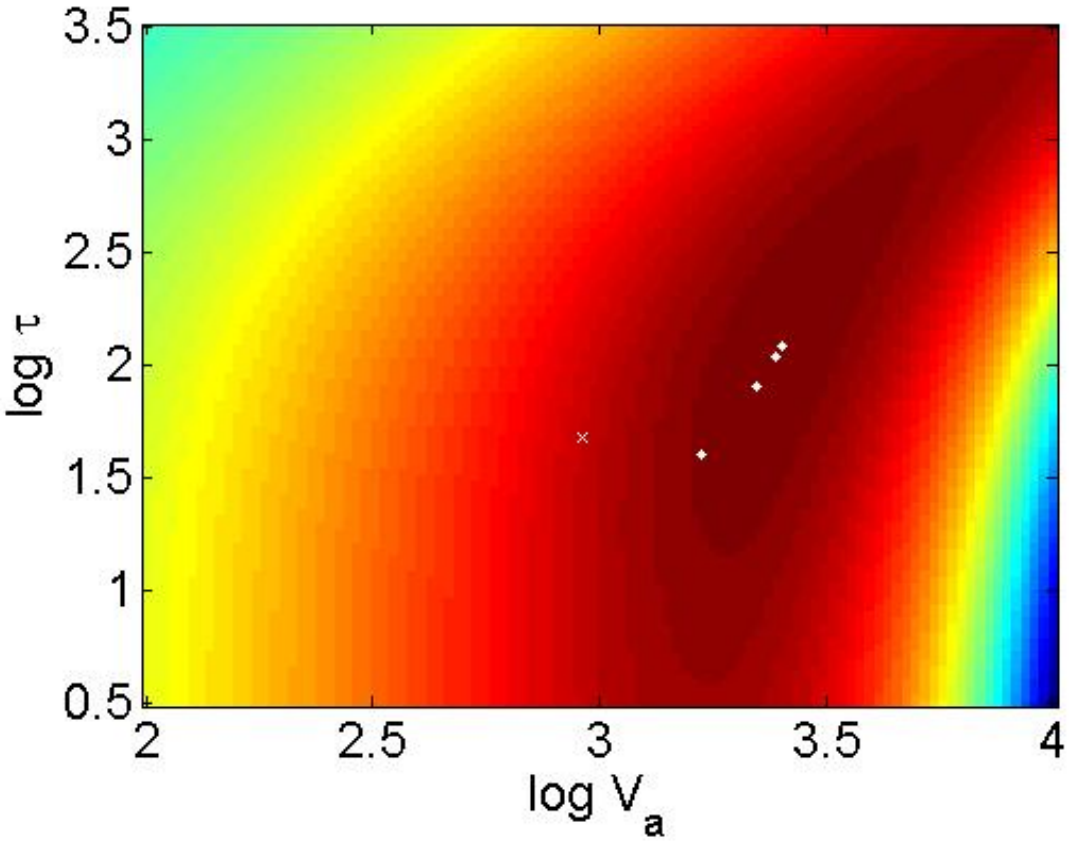
Posterior Landscape

A plot of $\log[p(y|w)p(w)]$



VL optimisation

Path of 6 VL iterations (x marks start)



References

C. Bishop (2006) Pattern Recognition and Machine Learning, Springer.

G. Deco et al. (2008) The Dynamic Brain: From Spiking Neurons to Neural Masses and Cortical Fields. PLoS CB 4(8), e1000092.

K. Friston et al. (2007) Variational Free Energy and the Laplace Approximation. Neuroimage 34(1), 220-234.

C. Moler and C Van Loan (2003). Nineteen Dubious Ways to Compute the Exponential of a Matrix. SIAM Review 45(1),801-836.

W. Penny (2006) Variational Bayes. In SPM Book, Elsevier.

O. David et al. (2006) Dynamic causal modelling of evoked responses in EEG and MEG. Neuroimage 30(4) 1255-1272.

F. Grimbert and O. Faugeras (2006) Bifurcation analysis of Jansen's Neural Mass Model. Neural Computation 18, 3052-3068.

B. Jansen and V. Rit (1995) Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. Biol. Cybern. 73, 275-283.

G. Strang (1988). Linear Algebra and its Applications. Brooks/Cole.

H. Wilson (1999) Spikes, decisions and actions: the dynamical foundations of neuroscience. Oxford.

W. Penny (2011) Bayesian Inference, Dynamical Systems and the Brain. Lecture Notes, Homepage.