Brain Imaging RSG - Problem Formulation

Don Praveen Amarasinghe, Andrew Lam, Pravin Madhavan

Mathematics and Statistics Centre for Doctoral Training
University of Warwick

7th February 2011

Mathematics and Statistics Centre for Doctoral Training

THE UNIVERSITY OF WARWICK

Contents

Contents

Functional Magnetic Resonance Imaging (fMRI)

fMRI tries to assess "Brain Activity" indirectly, through measurements of the blood flow and oxygenation in the brain.

Functional Magnetic Resonance Imaging (fMRI)

fMRI tries to assess "Brain Activity" indirectly, through measurements of the blood flow and oxygenation in the brain.

- MRI machine sends out a radio frequency pulse which generates a magnetic field.

Functional Magnetic Resonance Imaging (fMRI)

fMRI tries to assess "Brain Activity" indirectly, through measurements of the blood flow and oxygenation in the brain.

- MRI machine sends out a radio frequency pulse which generates a magnetic field.
- The energy of the field is enough to cause the spin of protons in blood Haemoglobin molecules to change.

Functional Magnetic Resonance Imaging (fMRI)

fMRI tries to assess "Brain Activity" indirectly, through measurements of the blood flow and oxygenation in the brain.

- MRI machine sends out a radio frequency pulse which generates a magnetic field.
- The energy of the field is enough to cause the spin of protons in blood Haemoglobin molecules to change.
- Protons in oxygenated haemoglobin behave differently to deoxygenated haemoglobin.

Functional Magnetic Resonance Imaging (fMRI)

fMRI tries to assess "Brain Activity" indirectly, through measurements of the blood flow and oxygenation in the brain.

- MRI machine sends out a radio frequency pulse which generates a magnetic field.
- The energy of the field is enough to cause the spin of protons in blood Haemoglobin molecules to change.
- Protons in oxygenated haemoglobin behave differently to deoxygenated haemoglobin.
- When the pulse is turned off, the energy absorbed by the resonating protons is released.

Electroencephalography (EEG)

EEG directly monitors electrical activity in the brain.

Electroencephalography (EEG)

EEG directly monitors electrical activity in the brain.

- Numerous electrodes are placed on the scalp.

Electroencephalography (EEG)

EEG directly monitors electrical activity in the brain.

- Numerous electrodes are placed on the scalp.
- Each electrode detects a change in electrical potential at that point on the scalp.

Electroencephalography (EEG)

EEG directly monitors electrical activity in the brain.

- Numerous electrodes are placed on the scalp.
- Each electrode detects a change in electrical potential at that point on the scalp.
- Voltages between electrodes can then be used to chart the electrical activity inside the brain.

Limitations

We would like to address the following issues:

Limitations

We would like to address the following issues:

- Spatial Resolution - EEG can't pinpoint the location of neural activity.

Limitations

We would like to address the following issues:

- Spatial Resolution - EEG can't pinpoint the location of neural activity.
- Signal Noise - In both fMRI and EEG, there are issues of noise introduced through the detection process. The signal can even "disappear"!

Limitations

We would like to address the following issues:

- Spatial Resolution - EEG can't pinpoint the location of neural activity.
- Signal Noise - In both fMRI and EEG, there are issues of noise introduced through the detection process. The signal can even "disappear"!
- External validity - There is a time delay issue with fMRI. There are also problems in establishing a control reading to begin with.

The Task

The Task

- Create a time-indexed series of noisy images which mimic the motion of a signal

The Task

- Create a time-indexed series of noisy images which mimic the motion of a signal
- Apply a technique to help remove the noise from these images

The Task

- Create a time-indexed series of noisy images which mimic the motion of a signal
- Apply a technique to help remove the noise from these images
- Apply a technique to track the motion of the signal through these images

Contents

Model

Despite the technical difficulties with fMRI and EEG discussed previously, we seek to infer properties of the noisy signal.

Model

Despite the technical difficulties with fMRI and EEG discussed previously, we seek to infer properties of the noisy signal.
To do so, we look at sequence of brain images taken in time to trace brain activity associated with stimulus.

Model

Despite the technical difficulties with fMRI and EEG discussed previously, we seek to infer properties of the noisy signal.
To do so, we look at sequence of brain images taken in time to trace brain activity associated with stimulus. Two main objectives:

Model

Despite the technical difficulties with fMRI and EEG discussed previously, we seek to infer properties of the noisy signal.
To do so, we look at sequence of brain images taken in time to trace brain activity associated with stimulus. Two main objectives:

- Filter the noise out from the image taken at first time-point.

Model

Despite the technical difficulties with fMRI and EEG discussed previously, we seek to infer properties of the noisy signal.
To do so, we look at sequence of brain images taken in time to trace brain activity associated with stimulus. Two main objectives:

- Filter the noise out from the image taken at first time-point.
- The denoised data can be used to evolve the observed signal in time.

Original data is composed of noisy surfaces defined on the square domain $[-1,1] \times [-1,1]$.

Original data is composed of noisy surfaces defined on the square domain $[-1,1] \times [-1,1]$.
Model considers 2D function with rotational symmetry, given by

$$\phi(x,y) = \exp\left(-\beta((x-c_1)^2 + (y-c_2)^2)\right)$$

where β controls how spiked the signal is and $\mathrm{c} = (c_1, c_2)$ the location of the signal.

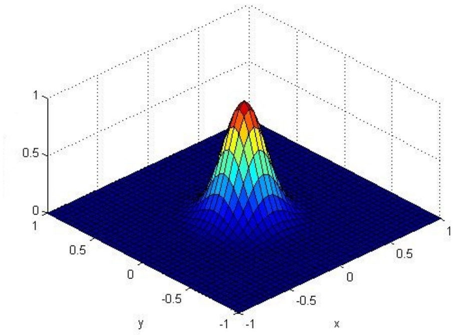

Figure: Plot of ϕ for $\beta = 20$ and $c = (0, 0)$

Add noise to the function by drawing independent samples from normal distribution with mean 0 and small variance and adding it to the function.

Add noise to the function by drawing independent samples from normal distribution with mean 0 and small variance and adding it to the function.

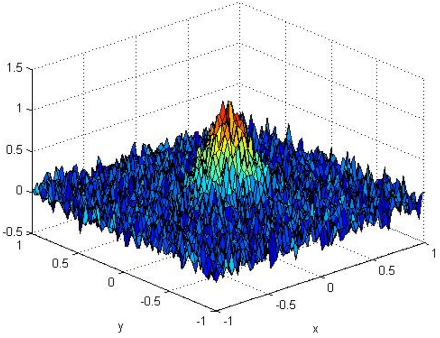

Figure: Plot of noisy signal for $\beta = 20$ and c $= (0, 0)$

We further improve the model by making β and c *noisy*.

We further improve the model by making β and c *noisy*.
β is a binary process taking two distinct values:

We further improve the model by making β and c *noisy*.

β is a binary process taking two distinct values:

- One, with low probability, which drowns the signal in the noise for a short period of time.

We further improve the model by making β and c *noisy*.

β is a binary process taking two distinct values:

- One, with low probability, which drowns the signal in the noise for a short period of time.
- The other, with high probability, in which the signal can be distinguished from the noise.

We further improve the model by making β and c *noisy*.

β is a binary process taking two distinct values:

- One, with low probability, which drowns the signal in the noise for a short period of time.
- The other, with high probability, in which the signal can be distinguished from the noise.

c follows a path of the form

$$c_2 = c_1^3 + u.$$

where c_1 moves from -1 to 1 and $u \sim \mathrm{Unif}([-0.1, 0.1])$.

(signal.avi)

Why have we chosen such models for these expressions?

Why have we chosen such models for these expressions?

- For β, want to incorporate key limitation of medical scanners, namely the disappearance of signal for short period of time.

Why have we chosen such models for these expressions?

- For β, want to incorporate key limitation of medical scanners, namely the disappearance of signal for short period of time.
- For c, want to capture the non-linear structure of the brain in order to characterise the signal more realistically.

Why have we chosen such models for these expressions?

- For β, want to incorporate key limitation of medical scanners, namely the disappearance of signal for short period of time.
- For c, want to capture the non-linear structure of the brain in order to characterise the signal more realistically. Regions of the brain activated by a stimulus need not lie on a path with simple geometry.

Why have we chosen such models for these expressions?

- For β, want to incorporate key limitation of medical scanners, namely the disappearance of signal for short period of time.
- For c, want to capture the non-linear structure of the brain in order to characterise the signal more realistically. Regions of the brain activated by a stimulus need not lie on a path with simple geometry.

Whilst our models do not fully reflect the complexity of such structures, it captures some of the non-linearity.

Contents

Contents

What are Bayes Factors?

What are Bayes Factors?

Suppose we have a null hypothesis

$$H_0 : \theta \in \Theta_0 \subset \Theta$$

which we want to test against an alternative hypothesis

$$H_1 : \theta \in \Theta \backslash \Theta_0$$

where Θ is the parameter space.

What are Bayes Factors?

Suppose we have a null hypothesis

$$H_0 : \theta \in \Theta_0 \subset \Theta$$

which we want to test against an alternative hypothesis

$$H_1 : \theta \in \Theta \backslash \Theta_0$$

where Θ is the parameter space.
The usual method of hypothesis testing involves a Likelihood Ratio Test Statistic, given by

$$S_{LR}(\mathbf{X}) = \frac{sup_{\Theta_0} L(\theta; \mathbf{X})}{sup_{\Theta} L(\theta; \mathbf{X})}$$

Under the Bayesian paradigm, we would like to modify this method to take into account our prior beliefs about the behaviour of the model. This gives rise to Bayes factors [Jeffreys (1935)].

Under the Bayesian paradigm, we would like to modify this method to take into account our prior beliefs about the behaviour of the model. This gives rise to Bayes factors [Jeffreys (1935)].
Bayes' Theorem says

$$\mathbb{P}(H_k|\mathbf{X}) = \frac{\mathbb{P}(\mathbf{X}|H_k)\mathbb{P}(H_k)}{\mathbb{P}(\mathbf{X}|H_0)\mathbb{P}(H_0) + \mathbb{P}(\mathbf{X}|H_1)\mathbb{P}(H_1)}$$

with $k = 0, 1$.

We then get

We then get

$$\frac{\mathbb{P}(H_0|\mathbf{X})}{\mathbb{P}(H_1|\mathbf{X})} = \frac{\mathbb{P}(\mathbf{X}|H_0)}{\mathbb{P}(\mathbf{X}|H_1)} \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)}$$

where

$$\mathbb{P}(\mathbf{X}|H_k) = \int \mathbb{P}(\mathbf{X}|\theta_k, H_k)\pi(\theta_k|H_k)d\theta_k$$

with θ_k the parameter under H_k with prior $\pi(\theta_k|H_k)$.
The highlighted term is the Bayes factor.

Why *Approximate* Bayes Factors?

The problem...

Why *Approximate* Bayes Factors?

The problem...

$$\mathbb{P}(\mathbf{X}|H_k) = \int \mathbb{P}(\mathbf{X}|\theta_k, H_k)\pi(\theta_k|H_k)d\theta_k$$

Unless we're lucky, we need to find ways of approximating this integral.

Why *Approximate* Bayes Factors?

The problem...

$$\mathbb{P}(\mathbf{X}|H_k) = \int \mathbb{P}(\mathbf{X}|\theta_k, H_k)\pi(\theta_k|H_k)d\theta_k$$

Unless we're lucky, we need to find ways of approximating this integral. There are various methods of doing this [Kass & Raftery (1995)]:

Why *Approximate* Bayes Factors?

The problem...

$$\mathbb{P}(\mathbf{X}|H_k) = \int \mathbb{P}(\mathbf{X}|\theta_k, H_k)\pi(\theta_k|H_k)d\theta_k$$

Unless we're lucky, we need to find ways of approximating this integral. There are various methods of doing this [Kass & Raftery (1995)]:

- Asymptotic Approximation
- Monte Carlo Methods
- MCMC & Metropolis-Hastings

Application - Image Segmentation

Application - Image Segmentation

We would like to use approximate Bayes factors to determine boundaries in a noisy image. In this particular example, we are interested in determining the number of gray levels to be used in an image.

Application - Image Segmentation

We would like to use approximate Bayes factors to determine boundaries in a noisy image. In this particular example, we are interested in determining the number of gray levels to be used in an image.

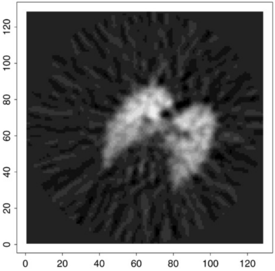

Figure: PET image of a dog's lung [Stanford & Raftery (2002)]

Approximate Bayes Factors

- We assume that the image has two “layers” (an actual image, and the observed image), giving rise to a Markov random field with the Potts Model.

- We assume that the image has two “layers” (an actual image, and the observed image), giving rise to a Markov random field with the Potts Model.
- We have a number of hypotheses, each representing a model using a different number of shades of grey (segments).

- We assume that the image has two "layers" (an actual image, and the observed image), giving rise to a Markov random field with the Potts Model.
- We have a number of hypotheses, each representing a model using a different number of shades of grey (segments).
- We use a Bayes factor approximation called the *Penalised Pseudolikelihood Criterion*, based upon maximum likelihood estimators, to compare favourability of these models (NB - Requires ICM first).

- We assume that the image has two “layers” (an actual image, and the observed image), giving rise to a Markov random field with the Potts Model.
- We have a number of hypotheses, each representing a model using a different number of shades of grey (segments).
- We use a Bayes factor approximation called the *Penalised Pseudolikelihood Criterion*, based upon maximum likelihood estimators, to compare favourability of these models (NB - Requires ICM first).
- Start with the model which has one shade of grey. Calculate the PLIC for that model, then move on to the next model. Iterate. Look out for a local maximum.

The result...

The result...

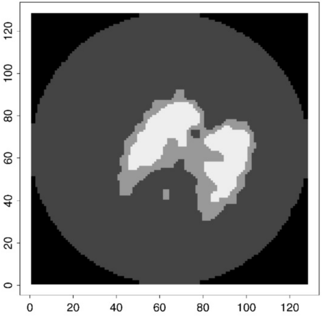

Figure: PET image of a dog's lung after final segmentation [Stanford & Raftery (2002)]

Contents

Short summary

Short summary

- A common tool for tracking problems/noise reduction is the Kalman filter.

Short summary

- A common tool for tracking problems/noise reduction is the Kalman filter.
- Given an observation X_t at time t, we want to infer on the state variable θ_t of a system. The state variables are linked to the observations via a matrix H.

Short summary

- A common tool for tracking problems/noise reduction is the Kalman filter.
- Given an observation X_t at time t, we want to infer on the state variable θ_t of a system. The state variables are linked to the observations via a matrix H.
- Measurements are typically noisy, so we include a noise term n_t.

Short summary

- A common tool for tracking problems/noise reduction is the Kalman filter.
- Given an observation X_t at time t, we want to infer on the state variable θ_t of a system. The state variables are linked to the observations via a matrix H.
- Measurements are typically noisy, so we include a noise term n_t.
- The Observation model is

$$X_t = H\theta_t + n_t.$$

Short summary

- A common tool for tracking problems/noise reduction is the Kalman filter.
- Given an observation X_t at time t, we want to infer on the state variable θ_t of a system. The state variables are linked to the observations via a matrix H.
- Measurements are typically noisy, so we include a noise term n_t.
- The Observation model is

$$X_t = H\theta_t + n_t.$$

- The state vector is updated by a transition matrix G with a noise process w_t,

$$\theta_t = G\theta_{t-1} + w_t.$$

Short summary

- We estimate θ_t with $\hat{\theta}_t$. Assume w_t and n_t are uncorrelated, with corresponding variance-covariance matrices Q and R.

Short summary

- We estimate θ_t with $\hat{\theta}_t$. Assume w_t and n_t are uncorrelated, with corresponding variance-covariance matrices Q and R.
- The simplest update of our estimate $\hat{\theta}_t$ is

$$\hat{\theta}_{t+1} = G\hat{\theta}_t.$$

Short summary

- We estimate θ_t with $\hat{\theta}_t$. Assume w_t and n_t are uncorrelated, with corresponding variance-covariance matrices Q and R.
- The simplest update of our estimate $\hat{\theta}_t$ is

$$\hat{\theta}_{t+1} = G\hat{\theta}_t.$$

- Denote the error $e_t = \theta_t - \hat{\theta}_t$ and its variance-covariance matrix P_t.

Short summary

- We estimate θ_t with $\hat{\theta}_t$. Assume w_t and n_t are uncorrelated, with corresponding variance-covariance matrices Q and R.
- The simplest update of our estimate $\hat{\theta}_t$ is

$$\hat{\theta}_{t+1} = G\hat{\theta}_t.$$

- Denote the error $e_t = \theta_t - \hat{\theta}_t$ and its variance-covariance matrix P_t.
- Assume the prior estimate of $\hat{\theta}_t$ is $\hat{\theta}_{t|t-1}$. The update equation, combining the old estimate and measurement, is

$$\hat{\theta}_t = \hat{\theta}_{t|t-1} + K_t(X_t - H\hat{\theta}_{t|t-1}),$$

where the Kalman gain K_t is derived while minimising the mean square error of the estimate.

Short summary

- We estimate θ_t with $\hat{\theta}_t$. Assume w_t and n_t are uncorrelated, with corresponding variance-covariance matrices Q and R.
- The simplest update of our estimate $\hat{\theta}_t$ is

$$\hat{\theta}_{t+1} = G\hat{\theta}_t.$$

- Denote the error $e_t = \theta_t - \hat{\theta}_t$ and its variance-covariance matrix P_t.
- Assume the prior estimate of $\hat{\theta}_t$ is $\hat{\theta}_{t|t-1}$. The update equation, combining the old estimate and measurement, is

$$\hat{\theta}_t = \hat{\theta}_{t|t-1} + K_t(X_t - H\hat{\theta}_{t|t-1}),$$

  where the Kalman gain K_t is derived while minimising the mean square error of the estimate.
- There is a similar update equation for P_t.

Applications

We highlight three uses of Kalman filter in the setting of medical images.

Applications

We highlight three uses of Kalman filter in the setting of medical images.

- EEG artifact removal

Applications

We highlight three uses of Kalman filter in the setting of medical images.

- EEG artifact removal
- EEG spike enhancement

Applications

We highlight three uses of Kalman filter in the setting of medical images.

- EEG artifact removal
- EEG spike enhancement
- Detecting activation regions

EEG artifact removal

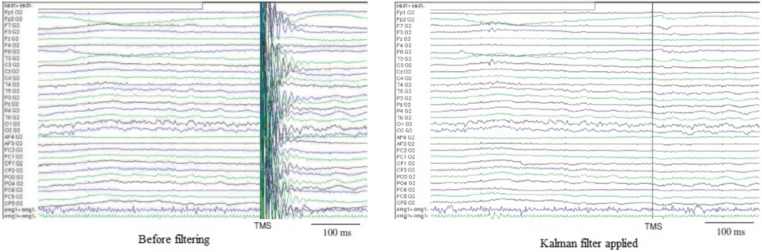

Figure: EEG artifact removal [Morbidi et al. (2007)]

EEG spike enhancement

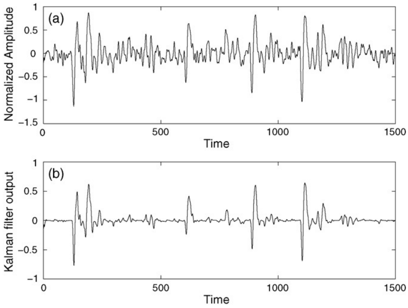

Figure: EEG spike enhancement [Oikonomou et al. (2006)]

Detecting activation regions

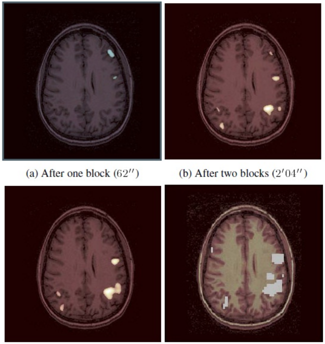
(a) After one block (62″) (b) After two blocks (2′04″)
(c) Final result (3′08″) (d) SPM'99 result

Figure: Incremental activation detection [Roche et al. (2004)]

Contents

Evaluation Metrics

Evaluation Metrics

- One key assumption needed to apply Kalman filters is that the noise is Gaussian. This may not necessarily be the case.

Evaluation Metrics

- One key assumption needed to apply Kalman filters is that the noise is Gaussian. This may not necessarily be the case.
- If we apply the Kalman filter as if noise was Gaussian, how would this affect the outcome of our analysis?

Evaluation Metrics

- One key assumption needed to apply Kalman filters is that the noise is Gaussian. This may not necessarily be the case.
- If we apply the Kalman filter as if noise was Gaussian, how would this affect the outcome of our analysis?
- We want to compare results that are derived from different models. We need some metric to evaluate this difference.

Evaluation Metrics

- One key assumption needed to apply Kalman filters is that the noise is Gaussian. This may not necessarily be the case.
- If we apply the Kalman filter as if noise was Gaussian, how would this affect the outcome of our analysis?
- We want to compare results that are derived from different models. We need some metric to evaluate this difference.
- We can use the matrix norm. But we want our metric to take into account the inherent stochasticity of the denoised data matrices.

Evaluation Metrics

- As an example, we use the discussed mathematical and statistical tools to generate a number of signal trajectory paths at every timepoint.

- As an example, we use the discussed mathematical and statistical tools to generate a number of signal trajectory paths at every timepoint.
- Then take the average of the computed paths and compare it with the true path.

- As an example, we use the discussed mathematical and statistical tools to generate a number of signal trajectory paths at every timepoint.
- Then take the average of the computed paths and compare it with the true path.

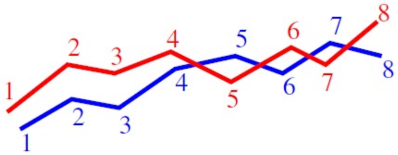

Figure: Example of a true path trajectory and the denoised + averaged one

- As an example, we use the discussed mathematical and statistical tools to generate a number of signal trajectory paths at every timepoint.
- Then take the average of the computed paths and compare it with the true path.

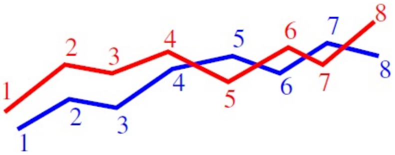

Figure: Example of a true path trajectory and the denoised + averaged one

- Various statistical metrics that compare such paths can be found in [Needham & Boyle, 2003]

Contents

Multiple signals

Multiple signals

- False positives arise from spatial delay or noise generated from the scanning process.

Multiple signals

- False positives arise from spatial delay or noise generated from the scanning process.
- There may also be spatial correlation among signals.

Multiple signals

- False positives arise from spatial delay or noise generated from the scanning process.
- There may also be spatial correlation among signals.
- Generate multimodal signal surfaces.

(ClusterSignal.avi)

Delayed detection

Delayed detection

- Temporal bias arises from detection process.

Delayed detection

- Temporal bias arises from detection process.
- What if a signal appears later in the time sequence?

Delayed detection

- Temporal bias arises from detection process.
- What if a signal appears later in the time sequence?
- Is this a delayed detection or just another false positive?

Delayed detection

- Temporal bias arises from detection process.
- What if a signal appears later in the time sequence?
- Is this a delayed detection or just another false positive?
- How would one set a threshold to decide that?

Delayed detection

- Temporal bias arises from detection process.
- What if a signal appears later in the time sequence?
- Is this a delayed detection or just another false positive?
- How would one set a threshold to decide that? – based on how often this signal appears in the time seqeuence?

Delayed detection

- Temporal bias arises from detection process.
- What if a signal appears later in the time sequence?
- Is this a delayed detection or just another false positive?
- How would one set a threshold to decide that? – based on how often this signal appears in the time seqeuence?
- Signals sometimes vanish from the trace – how would that change your threshold?

(DelaySignal.avi)

Difficulties and possible starting points

- Need to differentiate between the true signal and the false positives.

Difficulties and possible starting points

- Need to differentiate between the true signal and the false positives.
- Taking into account the correlation between signals.

Difficulties and possible starting points

- Need to differentiate between the true signal and the false positives.
- Taking into account the correlation between signals.
- The signal surface resembles a random field – a starting point would be to look at Random Field Theory.

Difficulties and possible starting points

- Need to differentiate between the true signal and the false positives.
- Taking into account the correlation between signals.
- The signal surface resembles a random field – a starting point would be to look at Random Field Theory.
- Apply thresholds to these surfaces and use hypothesis testing to locate activation regions.

Contents

Action Plan

- Generate noisy data – experiment with different parameter values to get a feel for how this toy model behaves. In addition, consider applying different noise distributions to your data. [1 day]

Action Plan

- Generate noisy data – experiment with different parameter values to get a feel for how this toy model behaves. In addition, consider applying different noise distributions to your data. [1 day]
- Read up on mathematical and statistical techniques which could be used to remove noise / track signals. [3 weeks]

Action Plan

- Generate noisy data – experiment with different parameter values to get a feel for how this toy model behaves. In addition, consider applying different noise distributions to your data. [1 day]
- Read up on mathematical and statistical techniques which could be used to remove noise / track signals. [3 weeks]
- Implement your chosen techniques – Test on dummy data before applying to the noisy data generated in the first step. [4 weeks]

Action Plan

- Compare your estimate the path of the signal with the actual data before noise was added to it. Furthermore, apply evaluation metrics to establish how sensitive your chosen techniques are to different noise distributions. [3 weeks]

Action Plan

- Compare your estimate the path of the signal with the actual data before noise was added to it. Furthermore, apply evaluation metrics to establish how sensitive your chosen techniques are to different noise distributions. [3 weeks]
- If you have time, consider applying the work you have done to the extension problems.

References

- [Jeffreys (1935)] – Some Tests of Significance, Treated by the Theory of Probability – Proceedings of the Cambridge Philosophy Society, Vol 31, 1935
- [Kass & Raftery (1995)] – Bayes Factors – Journal of the American Statistical Association, Vol 90, No 430, June 1995
- [Stanford & Raftery (2002)] – Approximate Bayes Factors for Image Segmentation: The Pseudolikelihood Information Criterion – IEEE Transactions on Pattern Analysis & Machine Intelligence, Vol 24, No 11, November 2002

References

- [Roche et al. (2004)] – Incremental Activation Detection in FMRI Series Using Kalman Filtering – IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2004, Vol 1, April 2004
- [Oikonomou et al. (2007)] – A Kalman Filter Based Methodology for EEG Spike Enhancement & Computer methods and programs – Biomedicine, Vol 85, No 2, Feburary 2007
- [Morbidi et al. (2007)] – A Kalman Filter Approach to Remove TMS-induced Artifacts from EEG Recordings – Proceedings of the European Control Conference, July 2007

References

- [Needham & Boyle (2003)] – Performance Evaluation Metrics and Statistics for Positional Tracker Evaluation – Computer Vision Systems, 2003

References

- [Needham & Boyle (2003)] – Performance Evaluation Metrics and Statistics for Positional Tracker Evaluation – Computer Vision Systems, 2003

Thank you for listening!