Optimization of Designs for fMRI

UCLA Advanced Neuroimaging Summer School

August 21, 2007

Thomas Liu, Ph.D.

UCSD Center for Functional MRI

Dictionary.com Unabridged (v 1.1) - Cite This Source

op·ti·mize [op-tuh-mahyz] Pronunciation Key - Show IPA Pronunciation *verb*, **-mized, -miz·ing.**

–verb (used with object)

1. to make as effective, perfect, or useful as possible.
2. to make the best of.
3. *Computers.* to write or rewrite (the instructions in a program) so as to maximize efficiency and speed in retrieval, storage, or execution.
4. *Mathematics.* to determine the maximum or minimum values of (a specified function that is subject to certain constraints).

–verb (used without object)

5. to be optimistic.

Also, *especially British*, **op·ti·mise.**

[Origin: 1835–45; OPTIM(UM) + -IZE]

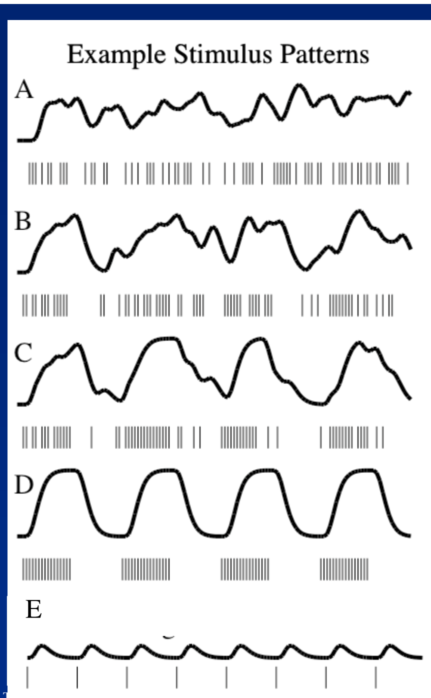
Why optimize?

- Scans are expensive.
- Subjects can be difficult to find.
- fMRI data are noisy
- A badly designed experiment is unlikely to yield publishable results.
- Time = Money

If your result needs a statistician then you should design a better experiment. *--Baron Ernest Rutherford*

What to optimize?

- Statistical Efficiency: maximize contrast of interest versus noise.
- Psychological factors: is the design too boring? Minimize anticipation, habituation, boredom, etc.



Which is the best design?

It depends on the experimental question.

Possible Questions

- Where is the activation?
- What does the hemodynamic response function (HRF) look like?

Model Assumptions

1) Assume we know the shape of the HRF but not its amplitude.
2) Assume we know nothing about the HRF (neither shape nor amplitude).
3) Assume we know something about the HRF (e.g. it's smooth).

General Linear Model

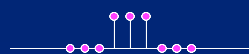
Data ↓ | Design Matrix ↓ | Nuisance Matrix ↓ | Additive Gaussian Noise ↓

$$\mathbf{y} = \mathbf{Xh} + \mathbf{Sb} + \mathbf{n}$$

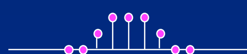
Parameters of Interest ↑ (h) | Nuisance Parameters ↑ (b)

Example 1: Assumed HRF shape

Stimulus



Convolve w/ HRF



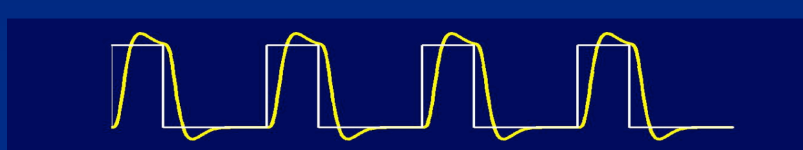
Design matrix depends on both stimulus and HRF

$$\mathbf{y} = \begin{bmatrix} 0 \\ 0 \\ 0.5 \\ 1 \\ 1 \\ 1 \\ 0.5 \\ 0 \\ 0 \end{bmatrix} h_1 + \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & 6 \\ 1 & 7 \\ 1 & 8 \\ 1 & 9 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} -1 \\ -2 \\ 0 \\ 3 \\ -1 \\ 1 \\ 2 \\ .5 \\ -.2 \end{bmatrix}$$

Parameter = amplitude of response

Design Regressor

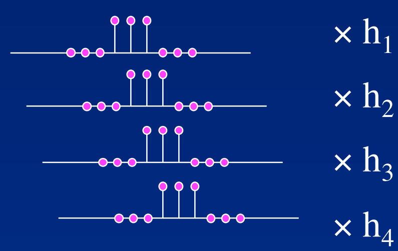
The process can be modeled by convolving the activity curve with a "hemodynamic response function" or HRF



Predicted response

Courtesy of FSL Group and Russ Poldrack

Example 2: Unknown HRF shape

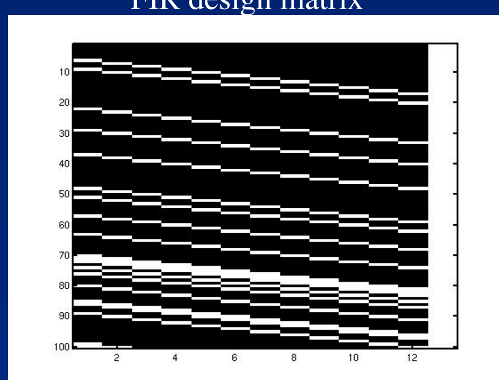


$$\mathbf{y} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & 6 \\ 1 & 7 \\ 1 & 8 \\ 1 & 9 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} -1 \\ -2 \\ 0 \\ 3 \\ -1 \\ 1 \\ 2 \\ .5 \\ -.2 \end{bmatrix}$$

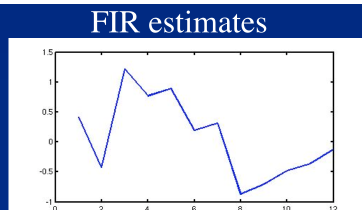
Note: Design matrix only depends on stimulus, not HRF

Unknown shape and amplitude

FIR design matrix



FIR estimates



Courtesy of Russ Poldrack

Test Statistic

Stimulus, neural activity, field strength, vascular state

$$t \propto \frac{\text{parameter estimate}}{\text{variance of parameter estimate}}$$

Thermal noise, physiological noise, low frequency drifts, motion

Also depends on Experimental Design!!!

Efficiency

$$\text{Efficiency} \propto \frac{1}{\text{Variance of Parameter Estimate}}$$

Efficiency

Example 1:

$$\text{Efficiency} \propto \frac{1}{\mathrm{Var}\left(\hat{\mathrm{h}}_1\right)}$$

Example 2:

$$\text{Efficiency} \propto \frac{1}{\mathrm{Var}\left(\hat{\mathrm{h}}_1\right) + \mathrm{Var}\left(\hat{\mathrm{h}}_2\right) + \mathrm{Var}\left(\hat{\mathrm{h}}_3\right) + \mathrm{Var}\left(\hat{\mathrm{h}}_4\right)}$$

Covariance Matrix

$$\mathrm{cov}\left(\hat{\mathbf{h}}\right) = \begin{bmatrix} \mathrm{var}\left(\hat{h}_1\right) & \mathrm{cov}\left(\hat{h}_1,\hat{h}_2\right) & \cdots & \mathrm{cov}\left(\hat{h}_1,\hat{h}_N\right) \\ \mathrm{cov}\left(\hat{h}_2,\hat{h}_1\right) & \mathrm{var}\left(\hat{h}_2\right) & \cdots & \mathrm{cov}\left(\hat{h}_2,\hat{h}_N\right) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}\left(\hat{h}_N,\hat{h}_1\right) & \mathrm{cov}\left(\hat{h}_N,\hat{h}_2\right) & \cdots & \mathrm{var}\left(\hat{h}_N\right) \end{bmatrix}$$

$$Efficiency \propto \frac{1}{Trace\left[\mathrm{cov}\left(\hat{\mathbf{h}}\right)\right]}$$

Known as an A-optimal design

General Linear Model

$$\underset{\text{Data}}{\mathbf{y}} = \underset{\text{Design Matrix}}{\mathbf{X}}\underset{\text{Hemodynamic Response}}{\mathbf{h}} + \mathbf{n}$$

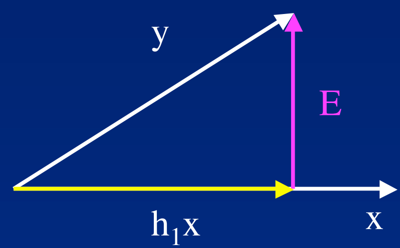
Nuisance Functions

Nuisance terms (constant term, linear drift, etc) are a fact of life in fMRI experiments.

However, to keep things simple, we will ignore the nuisance term Sb in the GLM for this talk.

The formulas we derive have the same form when nuisance terms are considered. Just replace X by $X_{\perp}$, where $X_{\perp}$ is obtained by projecting the nuisance terms out of the columns of X. See Liu et al 2001 and Liu and Frank 2004 for more details.

Principle of Orthogonality



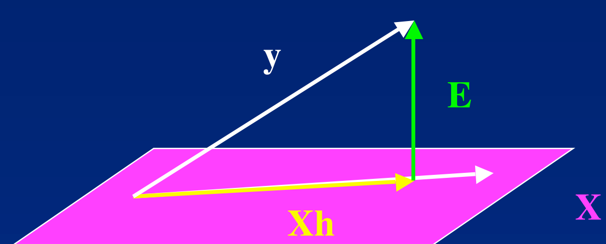
$$\mathrm{E}^T\mathbf{x} = 0$$

$$(\mathbf{y} - h_1\mathbf{x})^T\mathbf{x} = 0$$

$$h_1 = \frac{\mathbf{y}^T\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$$

Minimum error vector is orthogonal to the model space.

Principle of Orthogonality



$$\mathbf{X}^T\mathbf{E} = 0$$

$$\mathbf{X}^T(\mathbf{y} - \mathbf{Xh}) = \mathbf{0}$$

$$\Rightarrow \hat{\mathbf{h}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$$

Covariance of estimate

$$\begin{aligned}\mathrm{cov}(\hat{\mathbf{h}}) &= E\left(\left(\hat{\mathbf{h}} - E(\hat{\mathbf{h}})\right)\left(\hat{\mathbf{h}} - E(\hat{\mathbf{h}})\right)^T\right)\\ &= \left(\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T E\left((\mathbf{y}-\mathbf{X}\mathbf{h})(\mathbf{y}-\mathbf{X}\mathbf{h})^T\right)\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right)\\ &= \left(\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T E\left(\mathbf{n}\mathbf{n}^T\right)\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right)\\ &= \sigma^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\end{aligned}$$

Assume white noise for now

Depends on design

Depends on system, physiology, motion, etc.

Example 1: Assumed HRF shape

Assume we know the HDR shape $\mathbf{h}_0$ but not its amplitude h_1

$$\mathbf{h} = \mathbf{h}_0 h_1$$

GLM :

$$\begin{aligned}\mathbf{y} &= \mathbf{X}\mathbf{h} + \mathbf{n}\\ &= \mathbf{X}\mathbf{h}_0 h_1 + \mathbf{n}\\ &= \tilde{\mathbf{X}} h_1 + \mathbf{n} \quad \text{where } \tilde{\mathbf{X}} = \mathbf{X}\mathbf{h}_0\end{aligned}$$

Example 1: Assumed HRF shape

GLM :

$$\mathbf{y} = \tilde{\mathbf{X}} h_1 + \mathbf{n}$$

Efficiency :

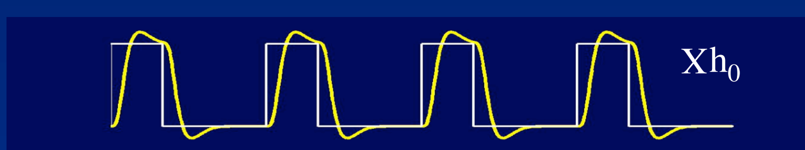
$$\xi \propto \frac{1}{\sigma^2 \operatorname{var}\left(\hat{h}_1\right)} = \frac{1}{\sigma^2 \left(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}}\right)^{-1}} = \frac{1}{\sigma^2 \left(\mathbf{h}_0^T \mathbf{X}^T \mathbf{X} \mathbf{h}_0\right)^{-1}} = \frac{\mathbf{h}_0^T \mathbf{X}^T \mathbf{X} \mathbf{h}_0}{\sigma^2}$$

Efficiency depends on both the design X and the assumed shape h_0 (plus intrinsic noise)

Interpretation

HRF

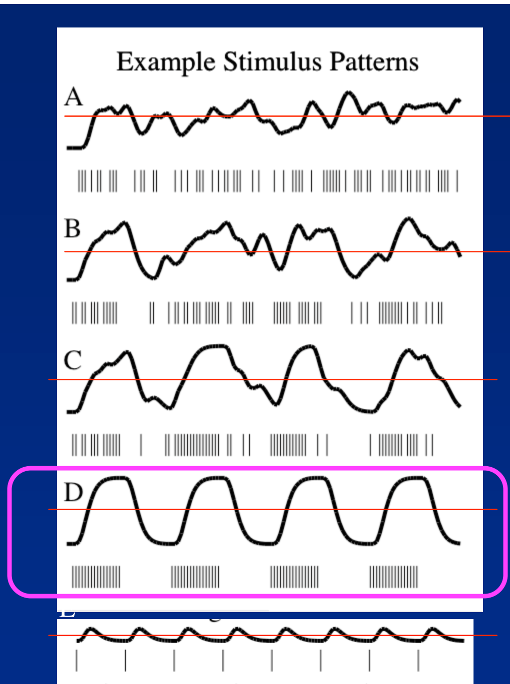
Predicted neural activity



Predicted response

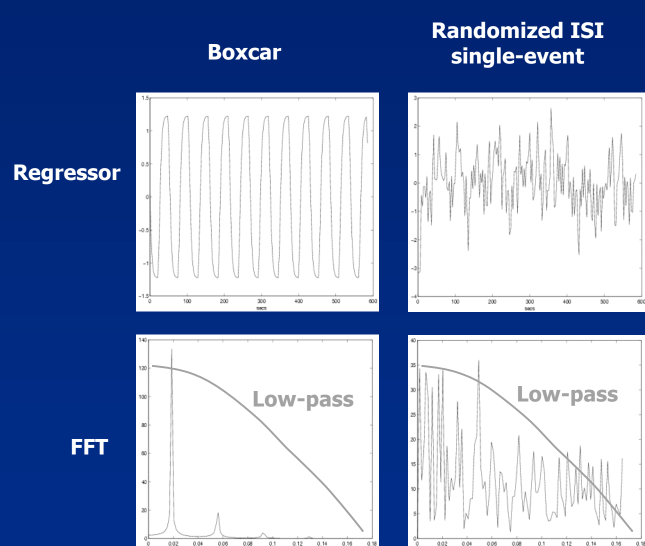
$$\mathbf{h}_0^T \mathbf{X}^T \mathbf{X} \mathbf{h}_0 = \left\|\mathbf{X}\mathbf{h}_0\right\|^2 \text{ -- measure of how "big" } \mathbf{X}\mathbf{h}_0 \text{ is}$$

 Modified from FSL Group and Russ Poldrack



Which design maximizes $\mathbf{h}_0^T \mathbf{X}^T \mathbf{X} \mathbf{h}_0$?

Frequency Domain Interpretation



Adapted From S. Smith and R. Poldrack

Example 2: Unknown HRF Shape

GLM :

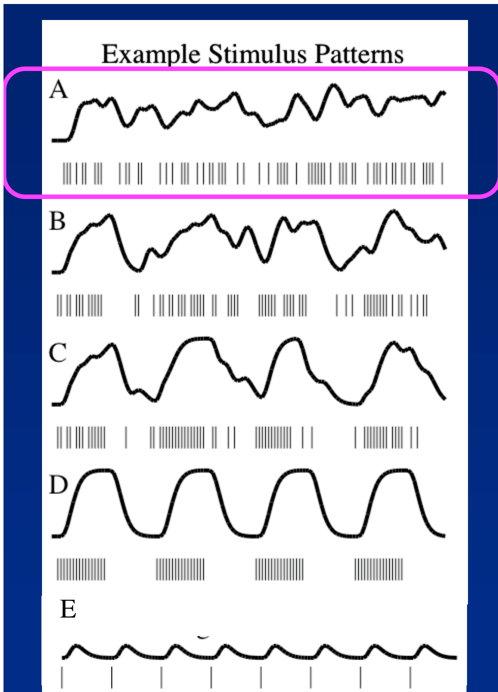
$$\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{n}$$

Efficiency :

$$\xi \propto \frac{1}{\sigma^2 \operatorname{var}(\hat{\mathbf{h}})}$$

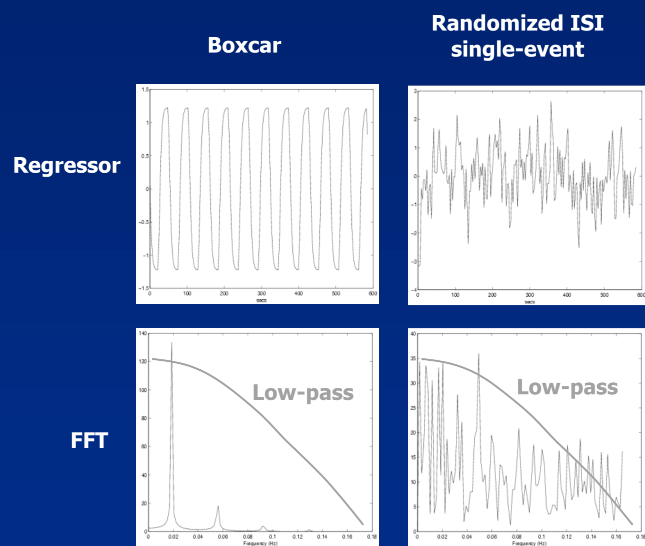
$$= \frac{1}{\sigma^2 Trace\left[\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right]}$$

Efficiency depends only on the design X (plus intrinsic noise)



Which design minimizes $\text{Trace}\left[\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right]$?

Frequency Domain Interpretation



Adapted From S. Smith and R. Poldrack

Maximizing Efficiency

$\text{Trace}\left[\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right]$ is minimized when the columns of X are orthogonal.

$\Rightarrow$ Shifted versions of the stimuli need to be orthogonal to each other

$\Rightarrow$ Shifted Randomized stimuli are orthogonal on average.

$\Rightarrow$ Shifted Block Designs are not orthogonal

Knowledge (Assumptions) about HRF

None — Some — Total

$$\xi \propto \frac{1}{\sigma^2 Trace\left[\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right]}$$

?

$$\xi \propto \frac{\mathbf{h}_0^T \mathbf{X}^T \mathbf{X} \mathbf{h}_0}{\sigma^2}$$

Depends only on X
Maximized by randomized designs

Depends on X and h_0
Maximized by block designs

Also referred to as detection power

Detection Power

When detection is the goal, we want to answer the question: is an activation present or not?

When trying to detect something, one needs to specify some knowledge about the "target".

In fMRI, the target is approximated by the convolution of the stimulus with the HRF.

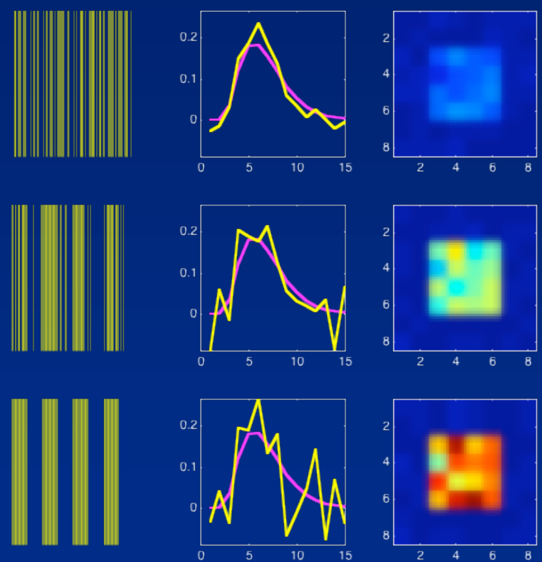
Once we have specified our target (e.g. stimuli and assumed HRF shape), the efficiency for estimating the amplitude of that target can be considered our detection power.

Power and Efficiency

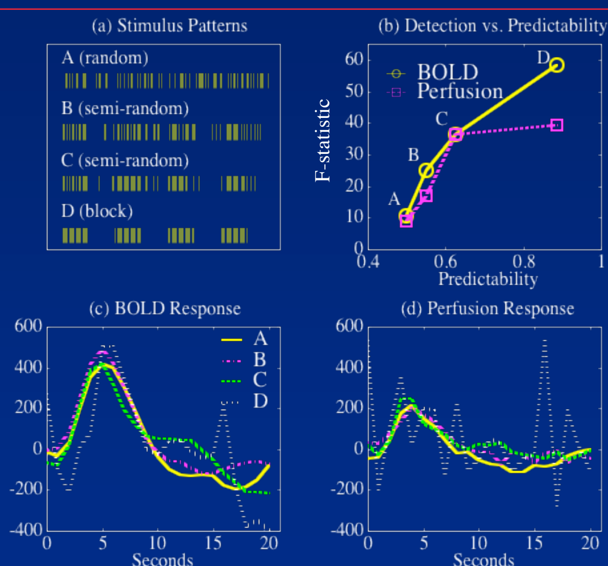
Random

SemiRandom

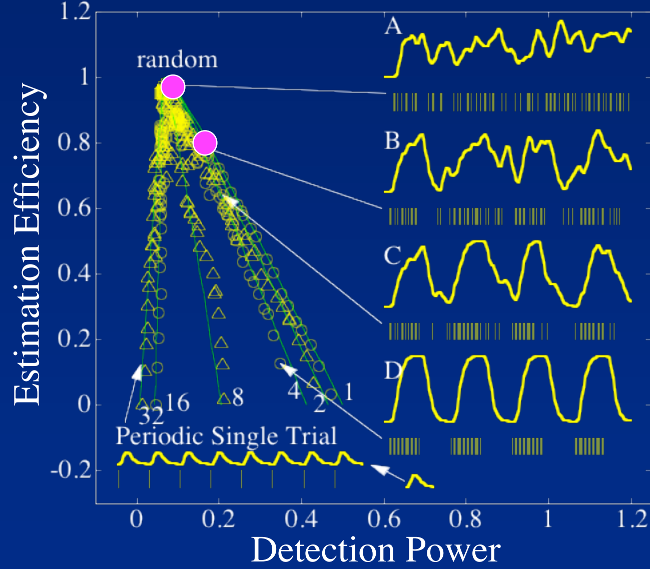
Block



Experimental Data

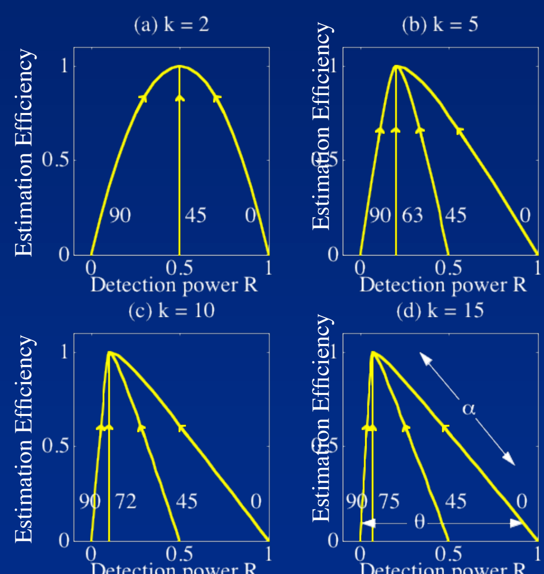


Efficiency vs. Power



T.T. Liu August 21, 2007

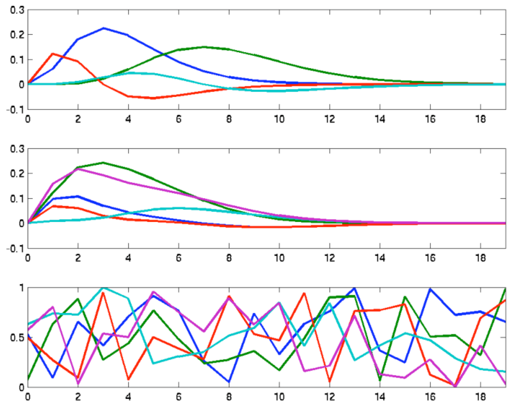
Theoretical Curves



T.T. Liu August 21, 2007

Basis Functions

If we know something about the shape, we can use a basis function expansion : $\mathbf{h} = \mathbf{Bc}$



4 basis functions

5 random HDRs using basis functions

5 random HDRs w/o basis functions

Here if we assume basis functions, we only need to estimate 4 parameters as opposed to 20.

Basis Functions

If we know something about the shape, we can use a basis function expansion : $\mathbf{h} = \mathbf{Bc}$

$$GLM : \mathbf{y} = \mathbf{XBc} + \mathbf{n} = \tilde{\mathbf{X}}\mathbf{c} + \mathbf{n}$$

$$Estimate : \hat{\mathbf{c}} = \left(\mathbf{B}^T \mathbf{X}^T \mathbf{X}\ \mathbf{B}\right)^{-1} \mathbf{B}^T \mathbf{X}^T \mathbf{y}$$

$$\hat{\mathbf{h}} = \mathbf{B}\hat{\mathbf{c}} = \mathbf{B}\left(\mathbf{B}^T \mathbf{X}^T \mathbf{X}\ \mathbf{B}\right)^{-1} \mathbf{B}^T \mathbf{X}^T \mathbf{y}$$

$$Efficiency : \xi = \frac{1}{\sigma^2 Trace\left[\mathbf{B}\left(\mathbf{B}^T \mathbf{X}^T \mathbf{X}\ \mathbf{B}\right)^{-1} \mathbf{B}^T\right]}$$

Efficiency now depends on both X and B

Knowledge (Assumptions) about HRF

None | Some | Total

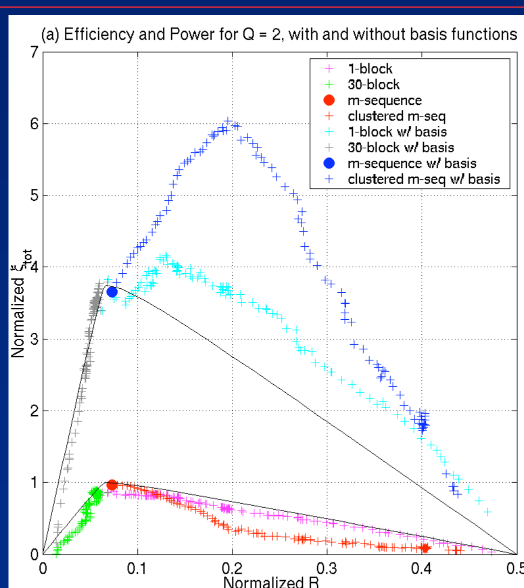
$$\xi \propto \frac{1}{\sigma^2 Trace\left[\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right]} \qquad \xi = \frac{1}{\sigma^2 Trace\left[\mathbf{B}\left(\mathbf{B}^T\mathbf{X}^T\mathbf{X}\ \mathbf{B}\right)^{-1}\mathbf{B}^T\right]} \qquad \xi \propto \frac{\mathbf{h}_0^T\mathbf{X}^T\mathbf{X}\mathbf{h}_0}{\sigma^2}$$

B = I

$B = h_0$

Depends on X and B
Maximized by semi-random designs.
Large increases in efficiency as compared to no assumptions

Efficiency with Basis Functions



Knowledge (Assumptions) about HRF

None — Some — Total

Experiments where you want to characterize in detail the shape of the HDR.

A reasonable compromise between 1 and 2. Detect activation when you sort of know the shape. Characterize the shape when you sort of know its properties

Experiments where you have a good guess as to the shape (either a canonical form or measured HDR) and want to detect activation.

Question

If block designs are so good for detecting activation, why bother using other types of designs?

Problems with habituation and anticipation

Less Predictable

Entropy

Perceived randomness of an experimental design is an important factor and can be critical for circumventing experimental confounds such as habituation and anticipation.

Conditional entropy is a measure of randomness in units of bits.

*R*th order conditional entropy (H_r) is the average number of binary (yes/no) questions required to determine the next trial type given knowledge of the *r* previous trial types.

2^{H_r} is a measure of the average number of possible outcomes.

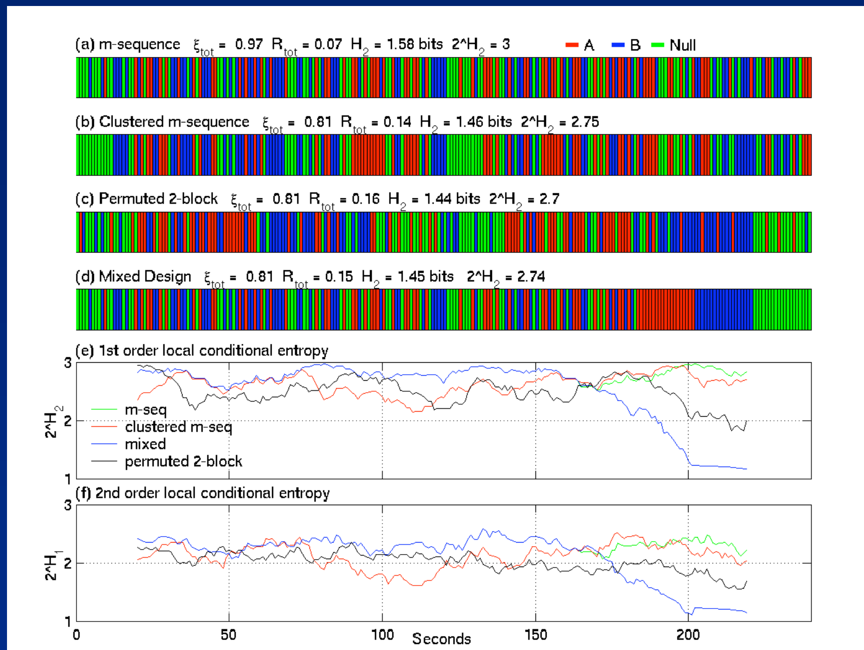
Entropy Example

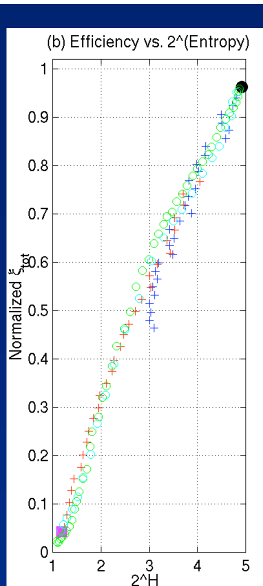
A A N A A N A A A N

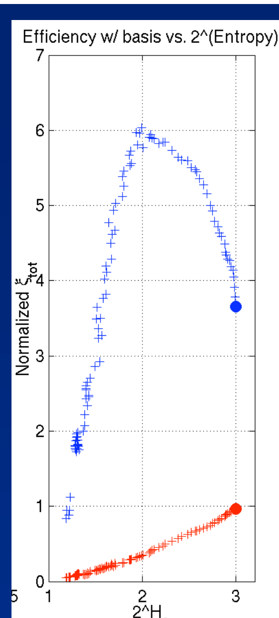
Maximum entropy is 1 bit, since at most one needs to only ask one question to determine what the next trial is (e.g. is the next trial A?). With maximum entropy, $2^1 = 2$ is the number of equally likely outcomes.

A C B N C B A A B C N A

Maximum entropy is 2 bits, since at most one would need to ask 2 questions to determine the next trial type. With maximum entropy, the number of equally likely outcomes to choose from is 4 (2^2).







Efficiency ∝ 2^(Entropy)

Multiple Trial Types

1 trial type + control (null)

A A N A A N A A A N

Extend to experiments with multiple trial types

A B A B N N A N B B A N A N A

B A D B A N D B C N D N B C N

Multiple Trial Types GLM

$$\mathbf{y} = \mathbf{Xh} + \mathbf{Sb} + \mathbf{n}$$

$$\mathbf{X} = [\mathbf{X}_1\ \mathbf{X}_2\ \ldots\ \mathbf{X}_Q]$$

$$\mathbf{h} = [\mathbf{h}_1^T\ \mathbf{h}_2^T\ \ldots\ \mathbf{h}_Q^T]^T$$

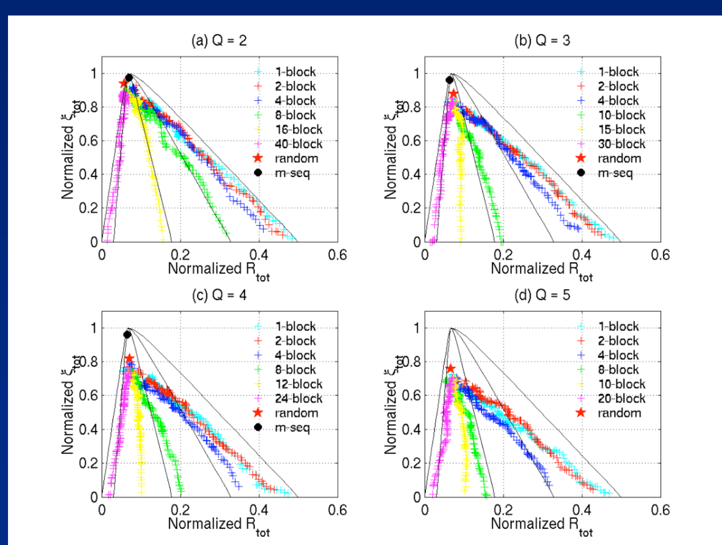
Multiple Trial Types Overview

Efficiency includes individual trials and also contrasts between trials.

$$R_{tot} = \frac{K}{\begin{pmatrix}\text{average variance of HRF amplitude estimates} \\ \text{for all trial types and pairwise contrasts}\end{pmatrix}}$$

$$\xi_{tot} = \frac{1}{\begin{pmatrix}\text{average variance of HRF estimates} \\ \text{for all trial types and pairwise contrasts}\end{pmatrix}}$$

Multiple Trial Types Trade-off



T.T. Liu August 21, 2007

Optimal Frequency

Can also weight how much you care about individual trials or contrasts. Or all trials versus events.
Optimal frequency of occurrence depends on weighting.
Example: With Q = 2 trial types, if only contrasts are of interest p = 0.5. If only trials are of interest, p = 0.2929.
If both trials and contrasts are of interest p = 1/3.

$$p = \frac{Q(2k_1 - 1) + Q^2(1 - k_1) + k_1^{1/2}\left(Q(2k_1 - 1) + Q^2(1 - k_1)\right)^{1/2}}{Q(Q-1)(k_1 Q - Q - k_1)}$$

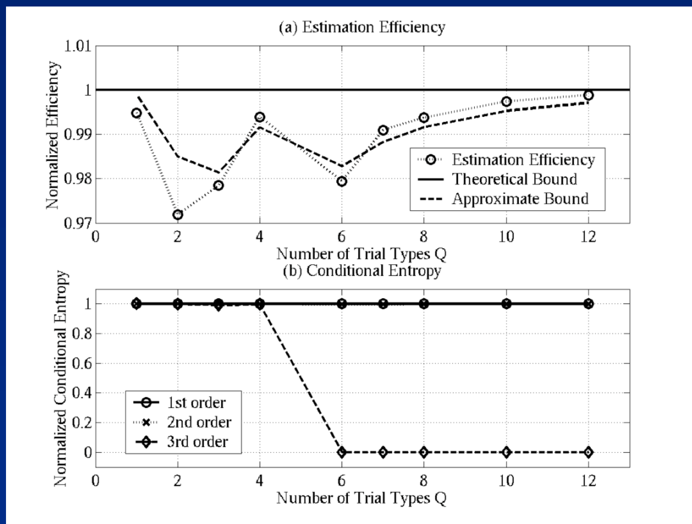
Design

As the number of trial types increases, it becomes more difficult to achieve the theoretical trade-offs. Random search becomes impractical.

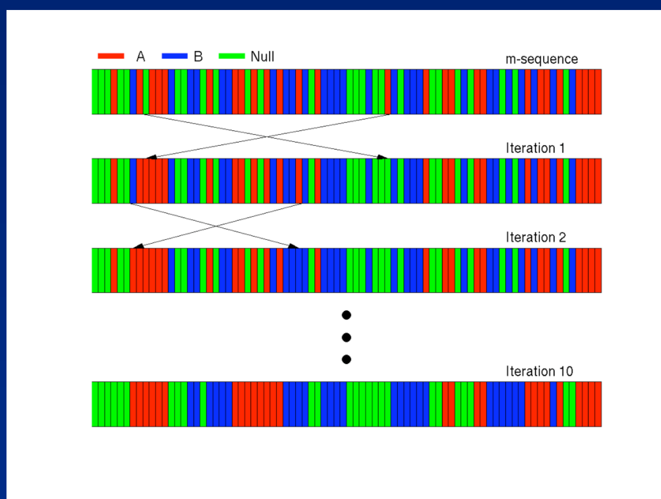
For unknown HDR, should use an m-sequence based design when possible.

Designs based on block or m-sequences are useful for obtaining intermediate trade-offs or for optimizing with basis functions or correlated noise.

Optimality of m-sequences



Clustered m-sequences



Topics we haven't covered.

The impact of correlated noise -- this will change the optimal design.

Impact of nonlinearities in the BOLD response.

Other optimization algorithms -- e.g. genetic algorithms.

Summary

- Efficiency as a metric of design performance.
- Efficiency depends on both experimental design and assumptions about HRF.
- Inherent tradeoff between power (detection of known HRF) and efficiency (estimation of HRF)