Optimization of Designs for fMRI
UCLA Advanced Neuroimaging Summer School
August 21, 2007

Thomas Liu, Ph.D.
UCSD Center for Functional MRI
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

\[ y = Xh + Sb + n \]

- **Data**
- **Design Matrix**
- **Nuisance Matrix**
- **Additive Gaussian Noise**

Parameters of Interest

Nuisance Parameters
Example 1: Assumed HRF shape

Stimulus

Convolve w/ HRF

Design matrix depends on both stimulus and HRF

Parameter = amplitude of response

Design Regressor

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

Design Regressor

Courtesy of FSL Group and Russ Poldrack
Example 2: Unknown HRF shape

\[
\begin{align*}
\times h_1 \\
\times h_2 \\
\times h_3 \\
\times h_4
\end{align*}
\]

Note: Design matrix only depends on stimulus, not HRF

Unknown shape and amplitude

\[
y = \begin{bmatrix}
0 & 0 & 0 & 0 & 1 & 1 & -1 \\
0 & 0 & 0 & 0 & 1 & 3 & 0 \\
1 & 0 & 0 & 0 & 1 & 4 & 3 \\
1 & 1 & 0 & 0 & 1 & 5 & h_1 \\
1 & 1 & 1 & 0 & 1 & 6 & h_2 \\
0 & 1 & 1 & 1 & 1 & 7 & 2 \\
0 & 0 & 1 & 1 & 1 & 8 & .5 \\
0 & 0 & 0 & 1 & 1 & 9 & -.2
\end{bmatrix}
\]

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

Example 1:
Efficiency $\propto \frac{1}{\text{Var}(\hat{h}_1)}$

Example 2:
Efficiency $\propto \frac{1}{\text{Var}(\hat{h}_1) + \text{Var}(\hat{h}_2) + \text{Var}(\hat{h}_3) + \text{Var}(\hat{h}_4)}$

Covariance Matrix

$\text{cov}(\hat{h}) = \begin{bmatrix}
\text{var}(\hat{h}_1) & \text{cov}(\hat{h}_1, \hat{h}_2) & \ldots & \text{cov}(\hat{h}_1, \hat{h}_N) \\
\text{cov}(\hat{h}_2, \hat{h}_1) & \text{var}(\hat{h}_2) & \ldots & \text{cov}(\hat{h}_2, \hat{h}_N) \\
\vdots & \vdots & \ddots & \vdots \\
\text{cov}(\hat{h}_N, \hat{h}_1) & \text{cov}(\hat{h}_N, \hat{h}_2) & \ldots & \text{var}(\hat{h}_N)
\end{bmatrix}$

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

Known as an A-optimal design
General Linear Model

\[ y = Xh + n \]

Data \hspace{1cm} \text{Design Matrix}

\[ n \]

\text{Hemodynamic Response}

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 \( S_b \) 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

Minimum error vector is orthogonal to the model space.

\[ E^T x = 0 \]
\[ (y - h_1 x)^T x = 0 \]
\[ h_1 = \frac{y^T x}{x^T x} \]

\[ X^T E = 0 \]
\[ X^T (y - Xh) = 0 \]
\[ \hat{h} = \left( X^T X \right)^{-1} X^T y \]
Covariance of estimate

\[ \text{cov}(\hat{h}) = E \left( (\hat{h} - E(\hat{h}))(\hat{h} - E(\hat{h}))^T \right) \]
\[ = \left( X^T X \right)^{-1} X^T E \left( (y - Xh)(y - Xh)^T \right) X \left( X^T X \right)^{-1} \]
\[ = \left( X^T X \right)^{-1} X^T E \left( nn^T \right) X \left( X^T X \right)^{-1} \]
\[ = \sigma^2 \left( X^T X \right)^{-1} \]

Assume white noise for now

Assume we know the HDR shape \( h_0 \) but not its amplitude \( h_1 \)

\[ h = h_0 h_1 \]

GLM:
\[ y = Xh + n \]
\[ = Xh_0 h_1 + n \]
\[ = \tilde{X} h_1 + n \quad \text{where} \quad \tilde{X} = Xh_0 \]
Example 1: Assumed HRF shape

GLM:
\[ y = \tilde{X}h_1 + n \]

Efficiency:
\[ \xi \propto \frac{1}{\sigma^2 \text{var}(h_1)} = \frac{1}{\sigma^2 (\tilde{X}^T \tilde{X})^{-1}} = \frac{1}{\sigma^2 (h_0^T X^T X h_0)^{-1}} = \frac{h_0^T X^T X h_0}{\sigma^2} \]

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

Interpretation

\[ h_0 \times X = Xh_0 \]

- Predicted neural activity
- Predicted response

\[ h_0^T X^T X h_0 = \|Xh_0\|^2 \]

- Measure of how "big" $Xh_0$ is
Which design maximizes $h_0^T X^T X h_0$?

Frequency Domain Interpretation

Adapted from S. Smith and R. Poldrack
Example 2: Unknown HRF Shape

**GLM:**
\[ y = Xh + n \]

**Efficiency:**
\[ \xi \propto \frac{1}{\sigma^2 \text{var}(h)} = \frac{1}{\sigma^2 \text{Trace}[(X^T X)^{-1}]} \]

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

---

Example Stimulus Patterns

Which design minimizes
\[ \text{Trace}[(X^T X)^{-1}] \]?
Maximizing Efficiency

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

⇒ Shifted versions of the stimuli need to be orthogonal to each other

⇒ Shifted Randomized stimuli are orthogonal on average.

⇒ Shifted Block Designs are not orthogonal
### Knowledge (Assumptions) about HRF

<table>
<thead>
<tr>
<th></th>
<th>None</th>
<th>Some</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\xi \propto \frac{1}{\sigma^2 \text{Trace}(X'^T X)^{-1}}$</td>
<td>?</td>
<td>$\xi \propto \frac{h_0^TX^TXh_0}{\sigma^2}$</td>
<td></td>
</tr>
<tr>
<td>Depends only on X</td>
<td>Depends on X and $h_0$</td>
<td>Maximized by randomized designs</td>
<td>Maximized by block designs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Also referred to as detection power</td>
<td></td>
</tr>
</tbody>
</table>

---

### 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

(a) Stimulus Patterns

A (random)
B (semi-random)
C (semi-random)
D (block)

(b) Detection vs. Predictability

(c) BOLD Response

(d) Perfusion Response
Basis Functions

If we know something about the shape, we can use a basis function expansion: \( h = Bc \)

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

\[
\text{Efficiency} = \frac{1}{\sigma^2 \text{Trace}\left[ B \left(B^T X^T B \right)^{-1} B^T \right]}
\]
Knowledge (Assumptions) about HRF

<table>
<thead>
<tr>
<th>None</th>
<th>Some</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\xi \propto \frac{1}{\sigma^2 \text{Trace} \left( X^T X \right)^{-1}}$</td>
<td>$\xi = \frac{1}{\sigma^2 \text{Trace} \left[ B \left( B^T X^T X B \right)^{-1} B^T \right]}$</td>
<td>$\xi \propto \frac{h_0^T X^T X h_0}{\sigma^2}$</td>
</tr>
</tbody>
</table>

B = I

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

B = $h_0$

Efficiency with Basis Functions

(a) Efficiency and Power for Q = 2, with and without basis functions
Knowledge (Assumptions) about HRF

None
Experiments where you want to characterize in detail the shape of the HDR.

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

Total
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.

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

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 A N A A N A A A N

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 $\propto 2^\text{Entropy}$
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

\[ y = Xh + Sb + n \]

\[ X = [X_1 \ X_2 \ \ldots \ X_Q] \]

\[ h = [h_1^T \ h_2^T \ \ldots \ h_Q^T]^T \]
Multiple Trial Types Overview

Efficiency includes individual trials and also contrasts between trials.

\[ R_{tot} = \frac{K}{\text{average variance of HRF amplitude estimates}} \]

\[ \xi_{tot} = \frac{1}{\text{average variance of HRF estimates}} \]

Multiple Trial Types Trade-off
**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 $p = 1/3$.

\[
p = \frac{Q(2k_1 - 1) + Q^2 (1 - k_1) + k_1^{1/2}(Q(2k_1 - 1) + Q^2 (1 - k_1))^{1/2}}{Q(Q - 1)(k_1Q - Q - k_1)}
\]

**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)